

On the statistical mechanics of structural vibration

R.S. Langley

Department of Engineering

University of Cambridge

Trumpington Street, Cambridge CB2 1PZ

rs121@eng.cam.ac.uk

Abstract

The analysis of the structural dynamics of a complex engineering structure has much in common with the subject of statistical mechanics. Both are concerned with the analysis of large systems in the presence of various sources of randomness, and both are concerned with the possibility of emergent laws that might be used to provide a simplified approach to the analysis of the system. The aim of the present work is to apply a number of the concepts of statistical mechanics to structural dynamic systems in order to provide new insights into the system behaviour under various conditions. The work is foundational, in that it is based on employing the fundamental equations of motion of the system in conjunction with various definitions of entropy, and no recourse is made to emergent laws that are accepted in thermodynamics. The analysis covers closed (undamped and unforced) and open (forced and damped) systems, linear and nonlinear systems, and both single systems and coupled systems. The fact that the system itself can be random leads to a number of results that differ from those found in classical statistical mechanics, where the initial conditions might be considered to be random but the Hamiltonian is taken to be well defined. For example, the occurrence of a stationary state in a closed system normally requires nonlinearity and coarse-graining of the statistical distribution, but neither condition is required for a random system. For coupled systems it is shown that under certain conditions both Statistical Energy Analysis (SEA) and Transient Statistical Energy Analysis (TSEA) are emergent laws, and insights are gained as to the validity of these laws. The analysis is supported by a number of numerical examples to illustrate key points.

Key words: vibration; random systems; entropy; statistical mechanics; energy flow

1. Introduction

The vibrational response of an engineering structure is a key design metric across many industrial sectors where there is a need to avoid excessive noise and vibration levels and to reduce the possibility of fatigue induced failure. Aircraft, automotive, and marine structures are all complex built-up systems, and the concern is often with frequencies at which the vibrational wavelength is short in comparison to the overall dimensions of the structure. This means that many degrees of freedom are needed to model the detailed deformation of the structure, and it is a difficult mathematical and computational task to formulate and solve the governing equations of motion. The general features of an engineering vibration analysis are illustrated in Figure 1: the system is excited from a set of initial conditions, and the aim is to predict the subsequent response. A complicating effect is that uncertainties can occur in all aspects of the analysis: the loading may be random, the initial conditions may be uncertain, and the system itself may have random properties due to manufacturing and material uncertainties. The task of the analyst is therefore to solve a large system of vibration equations within a probabilistic or otherwise uncertain framework, and hence there is a direct analogy with the problems encountered in the statistical mechanics of physical and chemical systems [1,2]. As discussed in what follows, the overall aim of the present work is to explore and exploit the connections between structural dynamics and statistical mechanics.

A simplified outline of the research agenda of statistical mechanics is shown in Figure 2. The subject is concerned with the dynamics of very large physical systems (1), and the standard approach is to adopt a probabilistic description of the response by introducing randomness in one of several possible ways (2). The equations governing the statistical response of the system can be formulated (3) but rarely solved in detail given their inherent complexity. As an alternative to a full solution, the properties and behaviour of global quantities such as entropy (of various types) and energy can be explored (4), with the aim of establishing relatively simple emergent laws that govern averaged properties of the system (5). Within this framework a conventional structural dynamics calculation might encompass stages (1) to (3), while classical thermodynamics is concerned with large scale laws of the type considered in stage (5) but arising directly from physical observations. A long term objective of statistical mechanics has been to provide a fundamental derivation of large scale laws, such as the Second Law of Thermodynamics, from the foundations of classical mechanics. It is fair to say that work on this objective has led to a great deal of controversy and has faced powerful counter-arguments in the form of the Loschmidt paradox regarding time reversibility and the Zermelo paradox regarding recurrence. Each of these paradoxes is

considered in detail later in the present work, and comprehensive reviews of the subject can be found in references [1-5]. A prevailing view is that for very large systems the Second Law holds almost always, although there is a finite (but exceedingly small) probability that the law may be violated.

In order to relate statistical mechanics to structural dynamics it is necessary to consider the concept of randomness that is adopted in each of the two subjects. In statistical mechanics the situation is not straight forward, as a range of views have been expressed by various authors. At one extreme Jaynes [6] considered the randomness in the formulation to be representative of a lack of knowledge, so that all probabilities should be interpreted in a subjective sense rather than an ensemble sense. In contrast, Gibbs [7], Boltzmann [8], and the Ehrenfests [9] were content to consider the statistical distribution to be defined across an ensemble of systems, presumably with random starting conditions causing the ensemble variation. Alternatively, Landau and Lifshitz [10] considered a single deterministic system and defined probability measures in terms of the fraction of time the system spends in particular states; this approach which can say little about non-stationary statistics [4] but corresponds to the ensemble approach in the stationary case if the system is ergodic. The subjective and ensemble views of uncertainty lead to the same mathematical equations and so the difference between the two approaches is philosophical rather than quantitative. In structural dynamics the situation is also multi-faceted in that uncertainties are classed as either aleatory or epistemic [11]; aleatory (or “irreducible”) uncertainties have ensemble (or frequentist) statistics, while epistemic uncertainties are “reducible” and have subjective probability distributions. Furthermore, non-probabilistic models of uncertainty such as interval or fuzzy descriptions are sometimes employed and these do not have a counterpart in statistical mechanics. If probabilistic descriptions are employed in structural dynamics then, mathematically at least, the subject is compatible with statistical mechanics, but with one proviso: in structural dynamics the system itself can be random, and this is not normally the case in classical statistical mechanics. It will be shown in the present work that this fact can be exploited to yield significant results.

The main aim of the present work is to extend structural dynamics from stages (1)-(3) of Figure 2 to stages (4) and (5) by considering entropy and related concepts. This is not simply a case of employing standard existing results, because structural dynamic systems differ considerably from those normally considered in statistical mechanics. Most notably: (i) the number of degrees of freedom is relatively small (for example 10 million degrees of freedom represents a very large structural dynamic system but a relatively small system in

terms of statistical mechanics), (ii) linear systems are of considerable practical importance, and so nonlinear and/or chaotic motions cannot be taken as a default condition, (iii) the system is often subjected to external environmental forcing and therefore open systems are of more practical interest than isolated systems, and (iv) as mentioned previously, the system itself can have random properties. It can be noted that an existing method that is employed in structural dynamics is related to stage (5) of Figure 2: a technique known as Statistical Energy Analysis (SEA) describes the energy flow in coupled systems in terms that are analogous to the heat conduction equation [12]. This method has previously been derived by a variety of approaches including modal energy flow [13], wave based methods [14,15], and thermal methods [16-19]. The novel thermal analysis presented in references [16] and [17] by Le Bot has close relations to the present work, although the approach taken therein does not start with the fundamental equations of dynamics but rather assumes a priori the key results of statistical mechanics including microstate equipartition [16] and the Maxwell-Boltzmann distribution [17]. Likewise, a very comprehensive analysis of the thermodynamics of dynamical systems is presented in reference [18] by Haddad et al, although it can be noted that the validity of the Second Law of Thermodynamics is assumed a priori as an axiom on p56 of that work, meaning that the work explores the consequences of the Second Law for dynamical systems rather than providing a fundamental derivation of the law. In contrast Carcaterra [19] provides an alternative view of the statistical mechanics of a closed dynamical system by employing the thermodynamic equations associated with the Khinchin entropy [20]; it is shown that the total entropy of a set of coupled oscillators will increase if there is a transition to a state of energy equipartition, but the increase need not be monotonic, and furthermore the state of equipartition need not be sustained, meaning that there may be to a subsequent decrease in entropy (a result which is consistent with the Zermelo paradox and in opposition to a strict interpretation of the Second Law). Taken together, references [16-19] are representative of previous work on the statistical mechanics of vibrating systems, and, as explained below, the present work differs significantly in approach.

The aim of the present work is to consider the transition from stages (1) to (5) of Figure 2 without assuming a priori any of the high level relations of statistical mechanics, thus providing a self-contained derivation of emergent laws such as SEA based only on the governing equations of motion. The key questions of practical interest that are addressed in what follows include: (i) does a closed structural dynamic system reach a state of statistical equilibrium, and if so, what is the nature of the equilibrium? (ii) what is the statistical nature

of the equilibrium state of an open system? (iii) can emergent laws such as SEA be derived using the methods of statistical mechanics, and to what extent does this shed light on the range of applicability of such laws?

Much of the work that follows is based on entropy, and so various definitions of entropy that have potential application to structural dynamics are considered in Section 2. The equations of motion of a deterministic dynamic system are then derived in Section 3, and by using the Fokker-Planck-Kolmogorov (FPK) equation a rate equation is found for the entropy, which includes the system damping and a term involving the forcing which takes the form of Fisher information [21]. For an undamped forced system it is found that the rate of change of entropy is equal to the Fisher information. The response of a closed system (unforced and undamped) is then considered in detail, and linear, integrable, and chaotic systems are each investigated. The behaviour of several types of entropy is explored and the extent to which “coarse graining” [9,22] leads to a stationary value of the entropy is considered. Attention is then turned to a forced-damped system, and the conditions under which the stationary response conforms to the statistical mechanical Maxwell-Boltzmann distribution are determined. The analysis is then extended in Section 4 to a system with random properties, and the questions of entropy, stationarity, and the Maxwell-Boltzmann distribution are re-addressed. In Sections 3 and 4 numerical results are presented for a number of example systems to illustrate key points of the analysis. Section 5 then considers two coupled subsystems, and both closed and forced-damped systems are analysed. The conditions under which emergent laws appear (most notably SEA) are considered, and it is shown that for closed systems the properties of the entropy play a key role. In contrast, for a forced-damped system it is shown that the properties of the Fisher information rather than the entropy lead to the emergence of the SEA equations. The conclusions of the work are then presented in Section 6.

2. Randomness and Vibrational Entropy

Consider a general N -degree-of-freedom vibrational system whose motion is described by a set of coordinates $\mathbf{x}(t)$, so that the instantaneous configuration of the system in phase space is defined by the position vector $(\mathbf{x}(t) \dot{\mathbf{x}}(t))$. If there is any uncertainty involved in the system dynamics then the position vector can be considered to be random, and the trajectory through phase space can be described by the joint probability density function (jpdf) $p(\mathbf{x}, \dot{\mathbf{x}}, t)$. The randomness in the system dynamics can arise from one or more of the following sources: (i)

uncertainty in the initial conditions, (ii) uncertainty in the system properties and the governing equations, and (iii) random excitation. Furthermore, the randomness might be interpreted in a frequentist sense or in a subjective (epistemic) sense; for example, one might consider multiple starts of the system, with the initial conditions drawn from an ensemble of possibilities, or one might consider a single start from initial conditions that are fixed but unknown. In the first case, the probability density function of the initial conditions measures frequency of occurrence across the ensemble, whereas in the second case the probability density function represents a subjective measure of belief in the various possibilities. As mentioned in the Introduction, the frequentist and the subjective views lead to identical mathematics, but they can result in different philosophical interpretations of the meaning of the analysis and controversy over the implications of the results. For example, the Gibbs approach to statistical mechanics [7] is based the jpdf of the positions and velocities of a set of particles, but it is not immediately clear how to interpret the implied randomness [22] – Gibbs suggested an ensemble of physical systems might be considered, while Jaynes [6] proposed that the theory relates to our state of knowledge rather than the actual world. Such controversies are avoided here by stating explicitly that a frequentist interpretation will be adopted, so a particular case is drawn from (i) an ensemble of possible starting positions, (ii) an ensemble of systems with varying properties, and (iii) an ensemble of possible random loading time histories. Although the analysis does not depend on this interpretation, it is helpful to think in engineering terms of, for example, an ensemble of random systems arising from a production line, or an ensemble of random environmental conditions. Much more could be said on this issue, and much has been written about the various interpretations of probability theory (for example [23]), but for present purposes it is sufficient to simply state the adopted view.

The jpdf $p(\mathbf{x}, \dot{\mathbf{x}}, t)$ contains a huge amount of detail regarding the dynamics of the system, and in some applications it is more useful to consider simpler metrics of the response, such as the energy or some other global quantity. One such metric is the entropy, which measures the degree of randomness of the response or, equivalently, the lack of information regarding the precise response. The standard measure of entropy in information theory is the Shannon entropy, which is defined as [22,24]

$$S(t) = - \int_R p(\mathbf{x}, \dot{\mathbf{x}}, t) \ln[p(\mathbf{x}, \dot{\mathbf{x}}, t)] d\mathbf{x} d\dot{\mathbf{x}}, \quad (1)$$

where R represents the accessible region of phase space. The Gibbs entropy that is employed in statistical mechanics is identical to the Shannon entropy, apart from the inclusion of a

multiplying factor k , the Boltzmann constant [7,22]. The properties of the entropy is a key topic of the present work, and for simplicity of notation the Shannon entropy rather than the Gibbs entropy will be employed. Another definition of entropy that is employed in statistical mechanics is the Boltzmann entropy, and this can be described by firstly considering the jpdf of a single degree of freedom of the complete system, say the j th coordinate:

$$p_j(x, \dot{x}, t) = - \int_{R'} p(x_1, x_2, \dots, x_{j-1}, x, x_{j+1}, \dots, x_N, \dot{x}_1, \dot{x}_2, \dots, \dot{x}_{j-1}, \dot{x}, \dot{x}_{j+1}, \dots, \dot{x}_N, t) d\mathbf{x}' d\dot{\mathbf{x}}', \quad (2)$$

where a prime indicates the omission of x_j and \dot{x}_j from the integral. The single particle distribution function $f(x, \dot{x}, t)$ is now defined as [1,22],

$$f(x, \dot{x}, t) = \sum_j p_j(x, \dot{x}, t), \quad (3)$$

with the physical interpretation that $f(x, \dot{x}, t) dx d\dot{x}$ is the average number of degrees of freedom whose position and velocity fall within the phase space area $dx d\dot{x}$ centred on the point (x, \dot{x}) . The Boltzmann H -function is defined in terms of this function as [1,22]

$$H(t) = \int_{R_2} f(x, \dot{x}, t) \ln[f(x, \dot{x}, t)] dx d\dot{x}, \quad (4)$$

where R_2 is the two-dimensional phase space region; the continuous Boltzmann entropy is then defined as [22]

$$S_H(t) = -kNH(t). \quad (5)$$

It can be noted that Eq. (5) differs from the well known Boltzmann combinatorial entropy [8,22], which will not be considered in the present work; the combinatorial entropy is concerned primarily with stationary states, while the present focus is on the evolution of a system through the governing equations of motion. A form of the continuous Boltzmann entropy will be employed here, but rather than use Eq. (5) it is more convenient to consider an entropy that coincides with the Shannon entropy under certain conditions. To this end, the “average” single degree of freedom jpdf is defined as

$$\bar{p}(x, \dot{x}, t) = (1/N) \sum_j p_j(x, \dot{x}, t), \quad (6)$$

and a Boltzmann-type entropy is defined as

$$S_B(t) = -N \int_{R_2} \bar{p}(x, \dot{x}, t) \ln[\bar{p}(x, \dot{x}, t)] dx d\dot{x}. \quad (7)$$

If the degrees of freedom are statistically independent and identically distributed, then Eqs. (1) and (7) yield the same entropy, and for this reason Eq. (7) is preferable to Eq. (5) for the

present purposes (the two entropies differ only by a multiplying factor and an additive constant).

A technique known as coarse graining is often employed in statistical mechanics [22], and this is based on dividing the phase space region R into M cells of equal volume. The probability density function within each cell is then replaced by the average value within the cell, leading to a probability density function that is piecewise constant across phase space, say $p_g(\mathbf{x}, \dot{\mathbf{x}}, t)$. This leads to coarse-grained values of the various entropies, so that the coarse-grained Shannon entropy is

$$S_g(t) = - \int_R p_g(\mathbf{x}, \dot{\mathbf{x}}, t) \ln[p_g(\mathbf{x}, \dot{\mathbf{x}}, t)] d\mathbf{x} d\dot{\mathbf{x}}, \quad (8)$$

and the coarse-grained Boltzmann-type entropy is

$$S_{B,g}(t) = -N \int_{R_2} \bar{p}_g(x, \dot{x}, t) \ln[\bar{p}_g(x, \dot{x}, t)] dx d\dot{x}. \quad (9)$$

As explained in what follows, it is possible to derive a number of inequalities between the various entropies which will be useful in the subsequent analysis.

The Boltzmann-type entropy given by Eq. (7) can be rewritten in the form

$$S_B(t) = - \sum_j \int_{R_2} p_j(x, \dot{x}, t) \ln[1 + (\bar{p} - p_j) / p_j] dx d\dot{x} - \sum_j \int_{R_2} p_j(x, \dot{x}, t) \ln[p_j(x, \dot{x}, t)] dx d\dot{x}. \quad (10)$$

Now the standard inequality $y \geq \ln(1 + y)$ can be applied to this expression, together with the fact that both \bar{p} and p_j integrate to unity over R_2 , to yield

$$S_B(t) \geq - \sum_j \int_{R_2} p_j(x, \dot{x}, t) \ln[p_j(x, \dot{x}, t)] dx d\dot{x}. \quad (11)$$

A similar manipulation of the Shannon entropy yields

$$\begin{aligned} S(t) &= \int_R p(\mathbf{x}, \dot{\mathbf{x}}, t) \ln[1 + (\hat{p} - p) / p] d\mathbf{x} d\dot{\mathbf{x}} - \int_R p(\mathbf{x}, \dot{\mathbf{x}}, t) \ln[\hat{p}] d\mathbf{x} d\dot{\mathbf{x}} \\ &\leq - \int_R p(\mathbf{x}, \dot{\mathbf{x}}, t) \ln[\hat{p}] d\mathbf{x} d\dot{\mathbf{x}} \end{aligned} \quad (12)$$

for any function \hat{p} that integrates to unity over R . Now taking the special case

$$\hat{p}(\mathbf{x}, \dot{\mathbf{x}}, t) = \prod_j p_j(x_j, \dot{x}_j, t), \quad (13)$$

yields

$$S(t) \leq - \sum_j \int_{R_2} p_j(x, \dot{x}, t) \ln[p_j(x, \dot{x}, t)] dx d\dot{x}. \quad (14)$$

It can be noted that equation (14) is a statement of the standard result that the Shannon entropy is always reduced by correlation between the variables [6]. Now equations (11) and (14) imply that the Boltzmann-type entropy is never less than the Shannon entropy:

$$S_B(t) \geq S(t). \quad (15)$$

A similar analysis can be applied to the coarse grained entropies to yield

$$S_{B,g}(t) \geq S_g(t). \quad (16)$$

A final pair of inequalities can be derived by noting that coarse graining will always increase the entropy: a uniform distribution has a higher entropy than a non-uniform distribution [6], and coarse graining replaces the actual distribution by a uniform distribution in each of the M cells. Thus

$$S_g(t) \geq S(t), \quad S_{B,g}(t) \geq S_B(t). \quad (17,18)$$

Having defined various measures of entropy, the properties of the entropy of a dynamic system will be explored in the following sections.

3. Single Deterministic Subsystem

3.1 General Equations

This section considers the dynamics of a system that is referred to as a “single subsystem”, meaning that the degrees of freedom are considered as a whole and not partitioned into a number of different groups (or subsystems). The system is taken to have deterministic properties (but may have random initial conditions) and the equations of motion are written as

$$\frac{d}{dt} \begin{pmatrix} \mathbf{x} \\ \dot{\mathbf{x}} \end{pmatrix} = \begin{pmatrix} \dot{\mathbf{x}} \\ -\mathbf{g}(\mathbf{x}, \dot{\mathbf{x}}) \end{pmatrix} + \begin{pmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{G} \end{pmatrix} \begin{pmatrix} \mathbf{0} \\ \mathbf{w}(t) \end{pmatrix}, \quad (19)$$

where $\mathbf{w}(t)$ is a vector of statistically independent Gaussian white noise processes, with each entry having $E[w_j(t)w_j(t+\tau)] = \delta(\tau)$. The vector $\mathbf{g}(\mathbf{x}, \dot{\mathbf{x}})$ represents the set of elastic and damping forces acting within the system, and the matrix \mathbf{G} governs the distribution of the external forces among the degrees of freedom. A system with non-white excitation can also be represented by Eq. (19) providing some of the degrees of freedom are non-physical and act to filter the white noise before application to the remaining degrees of freedom.

The response governed by Eq. (19) is a Markov vector process and it can be shown [25] that the jpdf of the state vector satisfies the Fokker-Planck-Kolmogorov (FPK) equation, which states

$$\frac{\partial p(\mathbf{x}, \dot{\mathbf{x}}, t)}{\partial t} = L_x[p(\mathbf{x}, \dot{\mathbf{x}}, t)], \quad (20)$$

where

$$L_x[p(\mathbf{x}, \dot{\mathbf{x}}, t)] = -\frac{\partial}{\partial x_j}[\dot{x}_j p(\mathbf{x}, \dot{\mathbf{x}}, t)] + \frac{\partial}{\partial \dot{x}_j}[g_j(\mathbf{x}, \dot{\mathbf{x}}) p(\mathbf{x}, \dot{\mathbf{x}}, t)] + \frac{1}{2} \frac{\partial^2}{\partial \dot{x}_j \partial \dot{x}_k}[B_{jk} p(\mathbf{x}, \dot{\mathbf{x}}, t)], \quad (21)$$

$$B_{jk} = G_{jr} G_{kr}. \quad (22)$$

The summation convention is employed in Eqs. (21) and (22), and in all that follows, so that a repeated index implies summation over that index.

Equation (21) can be used to derive an equation for the time evolution of the Shannon entropy of the system. Initially it can be noted from Eq. (1) that

$$\dot{S}(t) = -\int_R \{\ln[p(\mathbf{x}, \dot{\mathbf{x}}, t)] + 1\} \frac{\partial p(\mathbf{x}, \dot{\mathbf{x}}, t)}{\partial t} d\mathbf{x} d\dot{\mathbf{x}}, \quad (23)$$

and Eq. (20) then implies that

$$\dot{S}(t) = -\int_R \{\ln[p(\mathbf{x}, \dot{\mathbf{x}}, t)] + 1\} L_x[p(\mathbf{x}, \dot{\mathbf{x}}, t)] d\mathbf{x} d\dot{\mathbf{x}}. \quad (24)$$

If Eq. (21) is substituted into Eq. (24) then three terms appear on the right hand side of the equation. The first term can be evaluated to give

$$\int_R \{\ln[p(\mathbf{x}, \dot{\mathbf{x}}, t)] + 1\} \frac{\partial}{\partial x_j}[\dot{x}_j p(\mathbf{x}, \dot{\mathbf{x}}, t)] d\mathbf{x} d\dot{\mathbf{x}} = -\int_R \frac{\partial p(\mathbf{x}, \dot{\mathbf{x}}, t)}{\partial x_j} \dot{x}_j d\mathbf{x} d\dot{\mathbf{x}} = 0, \quad (25)$$

where it has been assumed that the jpdf is vanishingly small on the boundaries of R . The second term can be evaluated to give

$$\int_R \{\ln[p(\mathbf{x}, \dot{\mathbf{x}}, t)] + 1\} \frac{\partial}{\partial \dot{x}_j}[g_j(\mathbf{x}, \dot{\mathbf{x}}) p(\mathbf{x}, \dot{\mathbf{x}}, t)] d\mathbf{x} d\dot{\mathbf{x}} = \int_R \frac{\partial g_j(\mathbf{x}, \dot{\mathbf{x}})}{\partial \dot{x}_j} p(\mathbf{x}, \dot{\mathbf{x}}, t) d\mathbf{x} d\dot{\mathbf{x}}, \quad (26)$$

and the third term is

$$\int_R \{\ln[p(\mathbf{x}, \dot{\mathbf{x}}, t)] + 1\} \frac{1}{2} \frac{\partial^2}{\partial \dot{x}_j \partial \dot{x}_k}[B_{jk} p(\mathbf{x}, \dot{\mathbf{x}}, t)] d\mathbf{x} d\dot{\mathbf{x}} = -\frac{1}{2} \int_R \left\{ \frac{B_{jk}}{p(\mathbf{x}, \dot{\mathbf{x}}, t)} \frac{\partial p(\mathbf{x}, \dot{\mathbf{x}}, t)}{\partial \dot{x}_j} \frac{\partial p(\mathbf{x}, \dot{\mathbf{x}}, t)}{\partial \dot{x}_k} \right\} d\mathbf{x} d\dot{\mathbf{x}}. \quad (27)$$

It is now assumed that $\mathbf{g}(\mathbf{x}, \dot{\mathbf{x}})$ can be written as the sum of a set of conservative elastic forces and set of linear dissipative terms so that

$$g_j(\mathbf{x}, \dot{\mathbf{x}}) = f_j(\mathbf{x}) + C_{jr} \dot{x}_r, \quad (28)$$

in which case Eqs. (24)-(27) yield

$$\dot{S}(t) = -C_{jj} + F(t), \quad (29)$$

where

$$F(t) = \frac{1}{2} \int_R \left\{ \frac{B_{jk}}{p(\mathbf{x}, \dot{\mathbf{x}}, t)} \frac{\partial p(\mathbf{x}, \dot{\mathbf{x}}, t)}{\partial \dot{x}_j} \frac{\partial p(\mathbf{x}, \dot{\mathbf{x}}, t)}{\partial \dot{x}_k} \right\} d\mathbf{x} d\dot{\mathbf{x}}. \quad (30)$$

Any expression in the form of Eq. (30) has been termed “Fisher information” by Frieden [21], in this case with a weighting matrix \mathbf{B} . This terminology is not the conventional definition of the Fisher information, the difference being that the integrand in Eq. (30) involves differentiation with respect to the state vector, whereas the conventional expression involves differentiation with respect to parameters in the distribution [26]. It was shown in reference [27] that the Fisher information, as defined by Frieden, can be used to develop a probabilistic inference approach that differs from the principle of Maximum Entropy, and so, like entropy, the quantity is a useful measure of the amount of randomness in the system. A highly random system (with a broad jpdf) has a high value of entropy and a low value of Fisher information. In the absence of dissipation ($C_{jk} = 0$) Eq. (30) yields the curious result that the rate of change of entropy is equal to the Fisher information; equilibrium will be reached if and when the Fisher information is zero. Generally this will not occur – for example the case $N=1$ leads to classical Brownian motion with increasing variance [28] and increasing entropy. Another way to look at Eq. (29) in the general case ($C_{jk} \neq 0$) is that random excitation is a source of entropy and dissipation is a sink of entropy, which is a fairly obvious statement, although Eq. (29) expresses the fact in a mathematically succinct way.

The following subsections consider initially the entropy of an undamped unforced system, and attention is then turned to a damped forced system.

3.2 An undamped, unforced system

The entropy of an undamped unforced dynamic system has been the subject of a huge amount of research in statistical mechanics, ranging from the work of Boltzmann [8], Gibbs [7], and the Ehrenfests [9], to more recent discussions [3-5]. The key issue has been whether the Second Law of Thermodynamics (“the total entropy of an isolated system can never decrease over time”) can be derived from fundamental dynamical equations. It is fair to say that this is still a controversial topic, although some authors may feel that a sufficient understanding has been obtained of the matter [3]. Difficulties are posed by the paradoxes of Zermelo and Loschmidt [1]. The Zermelo paradox states that according to the Poincaré recurrence theorem any state of the system will recur (to any prescribed accuracy) an infinite number of

times – if the entropy increases from an initial state, it must therefore decrease at some later time when the initial state is recovered. The Loschmidt paradox is based on the fact that the equations of mechanics are reversible in time, and so increasing entropy must be accompanied by the symmetric possibility of decreasing entropy. Although much has been written on this topic, the present section addresses the issue from a mechanical vibrations viewpoint with the aim of providing a clear view of the issues involved and the conditions under which the Second Law might be expected to apply to the entropies defined in Section 1.

An immediate (and well known) result that follows from Eq. (29) is that in the absence of damping or forcing the Shannon entropy is constant; this entropy therefore follows the Second Law, albeit in a non-informative and trivial way. More interesting questions relate to whether the other entropy measures increase, and whether the response might achieve a stationary distribution. To address these issues it is helpful to consider initially an integrable Hamiltonian system. An integrable system is one in which the coordinates can be transformed into action-angle variables (corresponding to generalised momenta and displacements) by a canonical transformation. For any given starting condition the N action variables, \mathbf{J} say, are constant throughout the motion, and the N angle variables, $\boldsymbol{\theta}$ say, increase linearly with time, so that the rate is constant, say $\dot{\boldsymbol{\theta}} = \boldsymbol{\omega}(\mathbf{J})$. The response is said to be quasi-periodic, being a periodic function of each of the angle variables. A linear system is always integrable, and in this case the response can be expressed as a sum over the normal modes; the angular rates $\boldsymbol{\omega}$ correspond to the N natural frequencies, and the action variables can be related to the N amplitudes of the modal responses. In this case the angular rates are independent of the action variables, but in general this is not the case. A single degree of freedom Duffing oscillator [29], for example, is an integrable system, but the period of motion (related to the angular rate) is dependent on the amplitude of the motion (related to the action variable). Recurrence in an integrable system occurs when all of the angle variables $\boldsymbol{\theta}$ return (within a prescribed tolerance) to their starting values. If the angular rates are all rational fractions of the lowest rate, then recurrence can occur within a reasonably short time. It is more likely however that the angular rates are in irrational ratios, in which case the recurrence time can be extremely long, but finite. For example, Zhang and Liu have [30] considered the recurrence time of a chain of 15 masses connected by springs: it was found that each angle variable will return to within 0.3% of the starting value after 8.4×10^{31} cycles of the shortest natural period. Clearly this is an enormous duration, but the

important point is the philosophical significance of the existence of a recurrence time rather than the actual value of the time.

The fact that the angular rates of a linear system are independent of the action variables implies that whatever the starting conditions the system will always recur at the *same* time, t_r , say. This property is usually referred to as uniform recurrence [31]. Importantly this means that if the system is started randomly, with a prescribed jpdf, then the jpdf will recur at t_r , and hence *any* measure of the entropy will recur. If, for example, the Boltzmann-type entropy increases as the system vibrates, then it must reduce again as the time approaches the recurrence time. It is argued in references [3,9] (for example) that such reductions are likely to be very rare and of very short duration, but nonetheless they will exist. Furthermore, coarse graining will not resolve this issue since the coarse grained jpdf will also recur. These issues are illustrated in Figure 3 for the case of a set of 10 uncoupled linear oscillators, each having unit mass, with the natural frequencies arranged in a harmonic sequence so that $\omega_j = j$ rad/s. The system is taken to have a Gaussian jpdf at $t=0$, with all of the displacements and velocities being uncorrelated and of zero mean; the initial variance of each displacement is set at 1.0 m^2 , and all of the velocities have an initial variance of $0.2 \text{ m}^2/\text{s}^2$. This implies that the initial values of the Shannon and Boltzmann-type entropies are both equal to 20.33. The time history of four entropy measures is shown in the Figure: the Shannon entropy, the Boltzmann-type entropy, and coarse grained versions of these entropies obtained using a cell size of $0.389 \text{ m} \times 3.89 \text{ m/s}$. Because of the regularity of the natural frequencies the system has a very low recurrence time of 2π s, and because of the selected initial conditions the entropy has a lower recurrence time of π s and there is a symmetry (analogous to a time reversal) around the time $t = \pi / 2$ s. The results shown in the Figure highlight a number of the points raised in the foregoing discussion: (i) the Shannon entropy is constant, (ii) the Boltzmann-type entropy rises above the initial value but is subject to recurrence and therefore fluctuates, (iii) the coarse grained entropies are greater than the standard entropies, and each rises above the initial value, but the entropies are nonetheless subject to recurrence. It is clear that the Second Law of thermodynamics cannot be strictly true (in the simplistic sense of monotonically increasing entropy) for a deterministic linear system, even with random start conditions. This may sound like a radical statement to the vibration community, but the result is well known in statistical mechanics [3]. Attention is now turned to the case of a nonlinear integrable system.

If a nonlinear integrable system is given a single deterministic starting condition then the system will recur, just like a linear system, at a time determined by the angular rates. If however the system is given a random set of starting conditions, described by a jpdf, then the situation is more complicated. The random starting conditions will generally be associated with a distribution $p(\mathbf{J}, \boldsymbol{\theta})$ of the action-angle variables, and the angular rates $\boldsymbol{\omega}(\mathbf{J})$ corresponding to different action variables in this distribution will differ. Each point \mathbf{J} will therefore have a different recurrence time; for a continuous starting distribution there will be an infinite number of different recurrence times, and the time needed for the *whole* pdf to recur will be infinite (the system is said to have non-uniform recurrence [31]). The situation is therefore different to the linear case: the starting jpdf associated with a nonlinear system may not recur, and hence there is no reason why the starting entropy should recur. Furthermore, if the Boltzmann-type entropy and the Shannon entropy are the same at the starting time then Eq. (15) and the constancy of the Shannon entropy lead to the result

$$S_B(0) = S(0) \Rightarrow S_B(t) \geq S_B(0). \quad (31)$$

As noted below Eq. (7), the condition for the two entropies to coincide is that the degrees of freedom are uncorrelated and identically distributed; hence an initial distribution of this type will guarantee that the Boltzmann-type entropy will never decrease beyond the starting value. But this does not imply that the Boltzmann-type entropy can *never* decrease for a nonlinear system. Consider the case, for example, in which the system is started from the condition $S_B(0) = S(0)$ and then allowed to evolve for a time t , at which point the jpdf has become $p(\mathbf{x}, \dot{\mathbf{x}}, t)$. If the motion is now stopped and restarted with the jpdf $p(\mathbf{x}, -\dot{\mathbf{x}}, t)$ then the fact that the equations of motion are time reversible implies that the initial distribution will be recovered after a further time t . If the Boltzmann-type entropy increases between 0 and t , then it must decrease between t and $2t$. This implies that although the jpdf of a nonlinear system with random starting conditions can overcome the Zermelo paradox (recurrence), the Loschmidt paradox (time reversal) remains. Some starting distributions can lead to a decrease in the Boltzmann-type entropy, and even with a starting distribution that has $S_B(0) = S(0)$ there is no reason why the entropy should increase monotonically. Further insight can be obtained by approaching this problem from a different direction, rather than considering reversibility per se: given a starting distribution, does the system evolve to a stationary distribution, and is the entropy of this distribution greater than the initial entropy? This question is addressed in what follows.

As discussed above, if an integrable system is given an initial starting condition then during the subsequent motion the action variables \mathbf{J} will remain constant while the angle variables $\boldsymbol{\theta}$ will grow linearly with time. This motion is conventionally viewed as a trajectory on an N -torus [32], but for present purposes it is more convenient to think of the trajectory as a fixed point in \mathbf{J} -space and a linear trajectory in $\boldsymbol{\theta}$ -space. Because the angle variables are only uniquely defined modulo 2π , the $\boldsymbol{\theta}$ -space can be viewed as an N -dimensional hypercube of side length 2π ; a trajectory leaving the hypercube at say $\theta_j = 2\pi$ will re-enter the hypercube at $\theta_j = 0$, and the gradient of the trajectory is determined by the angular rates $\boldsymbol{\omega}(\mathbf{J})$. If the angular rates have rational ratios then the trajectory will repeat after a certain number of transits across the hypercube; if the angular rates have irrational ratios then the trajectory will never actually repeat, but (as discussed above) there will be a recurrence time after which the trajectory will fall arbitrarily close to the initial transit (more strictly, the recurrence time will depend on the tolerance placed on the proximity of the two transits). We can now consider the case in which the system is given a random start with an initial jpdf $p(\mathbf{J}, \boldsymbol{\theta})$ on the action-angle variables, which evolves to the jpdf $p(\mathbf{J}, \boldsymbol{\theta}, t)$ after a time t . The question was posed above as to whether $p(\mathbf{J}, \boldsymbol{\theta}, t)$ will tend to a stationary distribution at large times, and this can be addressed by considering the relation

$$p(\mathbf{J}, \boldsymbol{\theta}, t) = p(\boldsymbol{\theta}|\mathbf{J}, t)p(\mathbf{J}), \quad (32)$$

where the constancy of $p(\mathbf{J})$ has been noted. The initial conditional distribution of the angle variables can be visualised as a cloud in the $\boldsymbol{\theta}$ -hypercube. As time progresses, the cloud will propagate through the cube, but *without dispersion* since each point in the cloud has the same set of angular rates $\boldsymbol{\omega}(\mathbf{J})$ and therefore all of the trajectories will be parallel to each other. It is immediately apparent that $p(\boldsymbol{\theta}|\mathbf{J}, t)$ cannot become stationary, since the location of the cloud in the $\boldsymbol{\theta}$ -hypercube will vary with time for all time. It then follows that $p(\mathbf{J}, \boldsymbol{\theta}, t)$ cannot become stationary, and in principle the Boltzmann-type entropy $S_B(t)$ can vary for all time. The one exception to this conclusion is the case in which the initial distribution is uniform over the hypercube, in which case the distribution will not change with time and the starting condition is actually a stationary solution – in fact it is obvious that any distribution $p(\mathbf{J}, \boldsymbol{\theta})$ which depends on \mathbf{J} alone will be a stationary distribution. This special case is of limited interest however, since clearly all of the entropies will remain constant throughout the motion.

Given that in general $S_B(t)$ can fluctuate, it is of interest to consider whether the coarse-grained entropies $S_{B,g}(t)$ and $S_g(t)$ might nonetheless achieve stationary values. It was noted above that all the trajectories in $\boldsymbol{\theta}$ -space associated with $p(\boldsymbol{\theta}|\mathbf{J},t)$ will be parallel for a given value of \mathbf{J} , so that a cloud representing the distribution of the initial conditions will translate through space without dispersion. If we now consider the path of two clouds associated with different values of \mathbf{J} , then even if these clouds overlap at the initial time they will move away from each other as time progresses since they have different values of the angular rates $\boldsymbol{\omega}(\mathbf{J})$. At any late time t the clouds are likely to be in completely different positions in the $\boldsymbol{\theta}$ -hypercube. This process is illustrated schematically in Figure 4 for the case $N=2$. If this reasoning is extended to the clouds associated with a large number of different values of \mathbf{J} then it can be expected that eventually the clouds will together fill the hypercube; the amount of time needed for this spreading process to occur will depend on the proximity of the angular rates $\boldsymbol{\omega}(\mathbf{J})$ for the different values of \mathbf{J} , but no matter how close two sets of angular rates might be, the associated clouds are always likely to be in completely different positions after a sufficiently large time. It can therefore be postulated that if the jpdf $p(\mathbf{J},\boldsymbol{\theta},t)$ is averaged over \mathbf{J} by integrating over a small cell of volume V centred on some value \mathbf{J}_0 then the result will be

$$\langle p(\mathbf{J},\boldsymbol{\theta},t) \rangle = V^{-1} \int_V p(\boldsymbol{\theta}|\mathbf{J},t) p(\mathbf{J}) d\mathbf{J} \approx V^{-1} p(\mathbf{J}_0) \int_V p(\boldsymbol{\theta}|\mathbf{J},t) d\mathbf{J} \approx (2\pi)^{-N} p(\mathbf{J}_0), \quad (33)$$

where it has been assumed that $p(\boldsymbol{\theta}|\mathbf{J},t)$ averages over \mathbf{J} to a uniform distribution, for the reasons given above. Since coarse-graining involves averaging the jpdf over small regions of state-space, it follows from Eq. (33) that the coarse grained jpdf is likely to become (approximately) independent of $\boldsymbol{\theta}$ at late times and therefore stationary; furthermore, the form of the stationary jpdf will be entirely dependent on the initial distribution that is assigned to the action variables. The coarse grained entropies will also achieve stationary values, and given the constancy of the Shannon entropy and Eqs. (15)-(18) it follows that

$$S_g(0) = S(0) \Rightarrow S_g(\infty) \geq S_g(0), \quad (34)$$

$$S_{B,g}(0) = S(0) \Rightarrow S_{B,g}(\infty) \geq S_{B,g}(0). \quad (35)$$

The conditions on the left of Eqs. (34) and (35) will arise (see the discussion below Eq. (31)) if the degrees of freedom are initially uncorrelated and identically distributed, and if the initial distributions are smooth, so that the initial entropy is not affected by coarse-graining.

More generally, given Eq. (33), it can be expected that the final coarse-grained entropy (either Shannon or Boltzmann-type) will be greater than or equal to the initial value *regardless of the starting conditions*: the final distribution is uniform the θ variables, which is a maximum entropy condition. The fact that the system is time reversible does not prevent the occurrence of stationary coarse-grained entropy, since the above arguments hold for $t \rightarrow \pm\infty$ given an initial distribution at $t = 0$, and there is no claim that the entropy increases monotonically.

A simple example of the effect of coarse graining for a nonlinear system is afforded by a Duffing oscillator, which has a cubic restoring force and a vibration period which decreases with increasing energy of vibration. For a given initial condition the trajectory of the system in phase space is a simple closed orbit. If the system is given an initial jpdf that is uniform over a patch in phase space, not enclosing the origin, then the patch will spread with increasing time, since the points with high energy will circulate more quickly than those of low energy. In fact, as time progresses the patch will be stretched out into a spiral, lying between the highest energy orbit and the lowest energy orbit. The length of the spiral will continue to increase with increasing time, and neighbouring strands will get both thinner and closer together. The detailed pdf will never reach a stationary state, but at large times coarse graining will average over the strands of the spiral to produce a stationary jpdf and hence yield a coarse grained entropy that is time invariant. This behaviour is illustrated in Figures 5 and 6 for the example system $\ddot{x} + x^3 = 0$, with the initial jpdf taken to be a uniform distribution over the region $1 \leq x \leq 1.5$ m and $1 \leq \dot{x} \leq 1.5$ m/s. The evolution of this region with increasing time is shown in Figures 5(a-d), where the development of the spiral mentioned above can be observed (250,000 starting conditions within the initial rectangle were used to generate the plots). For a single degree of freedom system the Shannon and the Boltzmann-type entropy are equal and independent of time, but as shown in Figure 6 the coarse grained version of the entropy is not constant and increases to a stationary value which is greater than the initial value. The three curves shown in Figure 6 correspond to three different cell sizes used in the coarse graining procedure, each cell being square with a side length of either 0.05, 0.1, or 0.2. It can be seen that the stationary value of the coarse grained entropy is insensitive to the cell size over this range.

The forgoing discussion has considered the case of an integrable system. If the system is non-integrable and chaotic then it is much more difficult to reach general conclusions. Chaos can in some cases lead to strong mixing across phase space, so that a

starting jpdf will tend to fill the available phase space after a sufficiently long time [32]. However the Shannon entropy must be conserved, and in detail the distribution tends to fill the phase space in a filamentary way so that the jpdf is non-stationary, with continual motion of the filaments [32]. It can readily be imagined that in this case coarse graining will average out the filaments and lead to a stationary jpdf for which all of the entropies are greater than the starting values (similar to the Duffing oscillator considered above). Not all chaotic responses are so conceptually straight forward however. In some cases chaotic trajectories can become “trapped” in the vicinity of non-chaotic islands in phase space, meaning that these trajectories have extremely long recurrence times. The time taken for an initial jpdf to fill phase space can become so long that fluctuations from a stationary state can be large even at very late times. It is beyond the scope of the present work to consider such issues in detail, but reference [5] provides a very useful overview of the field. Mention can also be made of the field of ergodic systems [33], in which an alternative definition of entropy, the Kolmogorov-Sinai (KS) entropy, provides a measure of the spreading of trajectories.

As a final comment in this section, it can be concluded from the foregoing discussion that necessary conditions for a stationary distribution to arise in an undamped unforced system are *nonlinearity and coarse graining*. The (coarse grained) entropy of the stationary distribution will be greater than the starting value, although there is no reason why the entropy should be a monotonically increasing function of time. In the absence of a stationary distribution the Shannon entropy is constant but the Boltzmann-type entropy and the coarse grained entropies can fluctuate.

3.3 A damped, forced system

For a damped forced system the solution of the FPK equation, Eq. (21), will tend to a stationary limit at large times. Equation (29) then implies that the entropy decrease caused by damping will be balanced by the entropy increase caused by the excitation, so that

$$-C_{ij} + \frac{1}{2} \int_R \left\{ \frac{B_{jk}}{p_s(\mathbf{x}, \dot{\mathbf{x}})} \frac{\partial p_s(\mathbf{x}, \dot{\mathbf{x}})}{\partial \dot{x}_j} \frac{\partial p_s(\mathbf{x}, \dot{\mathbf{x}})}{\partial \dot{x}_k} \right\} d\mathbf{x} d\dot{\mathbf{x}} = 0, \quad (36)$$

where $p_s(\mathbf{x}, \dot{\mathbf{x}})$ is the stationary solution. The entropy of the final distribution may be greater or less than that of the initial distribution; the system is not closed, and so there is no reason to expect the entropy to increase. If the elastic forces in Eq. (28) are written in terms of a potential energy $U(\mathbf{x})$ so that

$$f_j(\mathbf{x}) = \frac{\partial U(\mathbf{x})}{\partial x_j}, \quad (37)$$

then it is of interest to consider the conditions under which the stationary solution has the form of the Maxwell-Boltzmann distribution that arises in statistical mechanics [1]

$$p_s(\mathbf{x}, \dot{\mathbf{x}}) = A \exp \left\{ -\alpha \left[(1/2) \dot{\mathbf{x}}_j \dot{\mathbf{x}}_j + U(\mathbf{x}) \right] \right\}, \quad (38)$$

for some constant α . It can be shown from Eq. (29) that the existence of a solution in the form of Eq. (38) requires

$$B_{jk} = \left(\frac{2}{\alpha} \right) C_{jk}, \quad (39)$$

so that there is a direct correspondence between the distribution of the excitation forces and the damping matrix. In detail, Eq. (39) states that the cross-spectral matrix of the excitation forces is proportional to the damping matrix. The occurrence of Eq. (38) will be of importance when considering coupled systems in Section 5.

It can be noted that the adoption of Eq. (37) tacitly involves the assumption that the system has a unit mass matrix, since more generally the equation would state $f_j = M_{jk}^{-1}(\partial U / \partial x_k)$, where \mathbf{M} is the mass matrix. If the mass matrix is not initially equal to the identity matrix then a change in coordinates can be performed to enforce this condition: for a linear system this could be a transformation to modal coordinates, while for a more general system a transformation of the type $\mathbf{x} = \mathbf{M}^{-1/2} \mathbf{y}$ could be applied, where \mathbf{M} is the mass matrix in the original set of coordinates \mathbf{y} . If this type of transformation is not performed, so that the mass matrix \mathbf{M} is not equal to the identity matrix, the Eq. (38) must be modified by the change $\dot{x}_j \dot{x}_j \rightarrow M_{jk} \dot{x}_j \dot{x}_k$. Equation (39) then remains a valid condition for the modified version of Eq. (38) to apply, with \mathbf{B} and \mathbf{C} representing respectively the cross-spectral forcing matrix and the damping matrix in the non-transformed coordinates. For ease of notation, a unit mass matrix will be adopted in all that follows.

All of the equations presented thus far relate to a system that is excited by white noise. In many practical engineering cases the concern is with the response of a multi-degree-of-freedom system subjected to band-limited excitation. As mentioned below Eq. (28) this situation can be addressed by augmenting the system with a filter; however this approach has the disadvantage of introducing non-physical degrees of freedom into the equations. For a linear system a simpler approach is to express the degrees of freedom in modal coordinates and then restrict the analysis to those modes that are resonant within the

excitation band. This is an approximation, in that it omits the response of the non-resonant modes, and furthermore it replaces the band-limited excitation of the resonant modes with white noise. Nonetheless, for light damping this can be expected to be a reasonable approach. For a nonlinear system a similar approach could be applied if the system can be represented as a nonlinear perturbation of an underlying linear system.

4. Single Random Subsystem

4.1 General equations

In this section the system under investigation is taken to have random properties which are described by a set of random parameters \mathbf{c} . These parameters may relate to the physical and/or geometrical properties of the system; for example in an automotive structure a spot weld may have uncertain stiffness properties, and a structural panel may have uncertain material properties. The FPK equation governing the response of the system, Eq. (20), remains valid for any individual member of the ensemble providing the following changes in notation are employed

$$p(\mathbf{x}, \dot{\mathbf{x}}, t) \rightarrow p(\mathbf{x}, \dot{\mathbf{x}}, t | \mathbf{c}), \quad \mathbf{g}(\mathbf{x}, \dot{\mathbf{x}}) \rightarrow \mathbf{g}(\mathbf{x}, \dot{\mathbf{x}}, \mathbf{c}), \quad \mathbf{B} \rightarrow \mathbf{B}(\mathbf{c}). \quad (40)-(42)$$

These changes simply note that the system equations depend on \mathbf{c} and that the jpdf governed by the FPK equation is conditional on the specified values of \mathbf{c} . If Eq. (20) is multiplied by the jpdf of the random parameters, $p(\mathbf{c})$ say, then it follows that

$$\frac{\partial p(\mathbf{x}, \dot{\mathbf{x}}, \mathbf{c}, t)}{\partial t} = L_x[p(\mathbf{x}, \dot{\mathbf{x}}, \mathbf{c}, t)]. \quad (43)$$

Integrating this result over the domain of the random parameters, R_c say, then yields

$$\frac{\partial p(\mathbf{x}, \dot{\mathbf{x}}, t)}{\partial t} = \int_{R_c} L_x[p(\mathbf{x}, \dot{\mathbf{x}}, \mathbf{c}, t)] d\mathbf{c}. \quad (44)$$

This equation can be recast in the form of a standard FPK equation in the unconditional jpdf of the response, so that

$$\frac{\partial p(\mathbf{x}, \dot{\mathbf{x}}, t)}{\partial t} = -\dot{x}_j \frac{\partial p(\mathbf{x}, \dot{\mathbf{x}}, t)}{\partial x_j} + \frac{\partial}{\partial \dot{x}_j} [\bar{g}_j(\mathbf{x}, \dot{\mathbf{x}}, t) p(\mathbf{x}, \dot{\mathbf{x}}, t)] + \frac{1}{2} \frac{\partial^2}{\partial \dot{x}_j \partial \dot{x}_k} \{ \bar{B}_{jk}(\mathbf{x}, \dot{\mathbf{x}}, t) p(\mathbf{x}, \dot{\mathbf{x}}, t) \}, \quad (45)$$

where

$$\bar{g}_j(\mathbf{x}, \dot{\mathbf{x}}, t) = \int_{R_c} g_j(\mathbf{x}, \dot{\mathbf{x}}, \mathbf{c}) p(\mathbf{c}, t | \mathbf{x}, \dot{\mathbf{x}}) d\mathbf{c}, \quad (46)$$

$$\bar{B}_{jk}(\mathbf{x}, \dot{\mathbf{x}}, t) = \int_{R_c} B_{jk}(\mathbf{c}) p(\mathbf{c}, t | \mathbf{x}, \dot{\mathbf{x}}) d\mathbf{c}. \quad (47)$$

To complete the formulation of the governing probabilistic equations, it is now assumed that each internal force $g_j(\mathbf{x}, \dot{\mathbf{x}}, \mathbf{c})$ can be expressed in terms of a random elastic component and a random dissipative component so that

$$g_j(\mathbf{x}, \dot{\mathbf{x}}, \mathbf{c}) = f_j(\mathbf{x}, \mathbf{c}) + C_{jr}(\mathbf{c})\dot{x}_r, \quad (48)$$

in which case Eq. (46) yields

$$\bar{g}_j(\mathbf{x}, \dot{\mathbf{x}}, t) = \int_{R_c} [f_j(\mathbf{x}, \mathbf{c}) + C_{jr}(\mathbf{c})\dot{x}_r] p(\mathbf{c}, t | \mathbf{x}, \dot{\mathbf{x}}) d\mathbf{c} = \bar{f}_j(\mathbf{x}, \dot{\mathbf{x}}, t) + \bar{C}_{jr}(\mathbf{x}, \dot{\mathbf{x}}, t)\dot{x}_r, \quad (49)$$

where \bar{f}_j and \bar{C}_{jk} are defined accordingly. Although Eq. (45) has the appearance of a standard FPK equation it should be noted that the coefficients in the equation, as defined by Eqs. (47) and (49), are dependent on the joint statistics of the random parameters and the response. Nonetheless, as will be seen in what follows, Eq. (45) is helpful in considering the time evolution of the entropy of the system.

The “full” Shannon entropy of the system can be defined as the entropy of the response in combination with the random parameters, so that

$$S_{\text{full}}(t) = - \int_{R+R_c} p(\mathbf{x}, \dot{\mathbf{x}}, \mathbf{c}, t) \ln[p(\mathbf{x}, \dot{\mathbf{x}}, \mathbf{c}, t)] d\mathbf{x} d\dot{\mathbf{x}} d\mathbf{c}. \quad (50)$$

It then follows from Eq. (43) that

$$\dot{S}_{\text{full}}(t) = - \int_{R+R_c} \{ \ln[p(\mathbf{x}, \dot{\mathbf{x}}, \mathbf{c}, t)] + 1 \} L_x[p(\mathbf{x}, \dot{\mathbf{x}}, \mathbf{c}, t)] d\mathbf{x} d\dot{\mathbf{x}} d\mathbf{c} = -E[C_{jj}] + F_{\text{full}}(t), \quad (51)$$

where $F_{\text{full}}(t)$ is given by Eq. (30) with the appropriate changes to introduce the random parameters \mathbf{c} into the arguments of the functions B_{jk} and p , and with integration over \mathbf{c} . The term $E[C_{jj}]$ is the unconditional expectation of the trace of the damping matrix, in contrast to the conditional expectation \bar{C}_{jk} . Now the Shannon entropy of the response alone can be defined as

$$S(t) = - \int_R p(\mathbf{x}, \dot{\mathbf{x}}, t) \ln[p(\mathbf{x}, \dot{\mathbf{x}}, t)] d\mathbf{x} d\dot{\mathbf{x}}, \quad (52)$$

and it follows from Eq. (45) that

$$\begin{aligned} \dot{S}(t) = & - \int_R p(\mathbf{x}, \dot{\mathbf{x}}, t) \frac{\partial}{\partial \dot{x}_j} \left\{ \bar{C}_{jk}(\mathbf{x}, \dot{\mathbf{x}}, t) \dot{x}_k \right\} d\mathbf{x} d\dot{\mathbf{x}} - \int_R p(\mathbf{x}, \dot{\mathbf{x}}, t) \frac{\partial \bar{f}_j(\mathbf{x}, \dot{\mathbf{x}}, t)}{\partial \dot{x}_j} d\mathbf{x} d\dot{\mathbf{x}} \\ & + \frac{1}{2} \int_R \frac{1}{p(\mathbf{x}, \dot{\mathbf{x}}, t)} \frac{\partial p(\mathbf{x}, \dot{\mathbf{x}}, t)}{\partial \dot{x}_j} \frac{\partial}{\partial \dot{x}_k} \left\{ \bar{B}_{jk}(\mathbf{x}, \dot{\mathbf{x}}, t) p(\mathbf{x}, \dot{\mathbf{x}}, t) \right\} d\mathbf{x} d\dot{\mathbf{x}}. \end{aligned} \quad (53)$$

The implications of Eqs. (51) and (53) will be explored in the following subsections for various cases.

4.2 An undamped, unforced system

For an undamped and unforced system it follows immediately from Eq. (51) that the “full” Shannon entropy is constant, since both terms on the right hand side of the equation are zero by definition. Thus

$$S_{\text{full}}(t) = S_{\text{full}}(0). \quad (54)$$

Furthermore, if the initial starting conditions are independent of the random parameters then

$$S_{\text{full}}(0) = S(0) + S_c(0), \quad (55)$$

where $S_c(t)$ is the Shannon entropy of the random parameters. As noted below Eq. (14), correlation between random variables always reduces the Shannon entropy, and hence the full Shannon entropy at some time t will be less than the entropy that would be obtained were the random parameters and the response uncorrelated at this time, i.e.

$$S_{\text{full}}(t) \leq S(t) + S_c(t). \quad (56)$$

Given the fact that the random parameters are independent of time, so that $S_c(t) = S_c(0)$, it follows from Eqs. (54)-(56) that

$$S(t) \geq S(0). \quad (57)$$

Thus, unlike a deterministic system, the Shannon entropy of a random system is not necessarily constant with time and can exceed an initial value (to be precise, the entropy in question here is the entropy of the response of an ensemble of random systems).

An insight into the fundamental differences between a random system and a deterministic system can be explored by considering the issues of reversibility and recurrence. A deterministic system is always reversible in the sense discussed in Section 3.2, i.e. the transition $p(\mathbf{x}, \dot{\mathbf{x}}, 0) \rightarrow p(\mathbf{x}, \dot{\mathbf{x}}, t)$ implies that $p(\mathbf{x}, -\dot{\mathbf{x}}, t)$ will return to the initial distribution in a time t . For a random system this property will clearly apply to the conditional jpdf $p(\mathbf{x}, \dot{\mathbf{x}}, t | \mathbf{c})$ but it will not apply to the unconditional jpdf $p(\mathbf{x}, \dot{\mathbf{x}}, t)$, simply

because $p(\mathbf{x}, -\dot{\mathbf{x}}, t)$ will not in general coincide with the probability distribution needed for reversal on any single realisation of the random system. Regarding recurrence, it can be noted that for a linear system the jpdf $p(\mathbf{x}, \dot{\mathbf{x}}, t | \mathbf{c})$ will be recurrent, but the recurrence time will depend upon the parameters \mathbf{c} ; given that there are an infinite number of members of the ensemble, each with a different recurrence time, the time required for simultaneous recurrence across the ensemble will be infinite, i.e. recurrence of $p(\mathbf{x}, \dot{\mathbf{x}}, t)$ will not occur. Given the absence of recurrence and reversibility, a random system avoids the paradoxes of Zermelo and Loschmidt, although there is no guarantee that the entropy will increase monotonically. It is, however, highly likely that a stationary distribution will be achieved at large times, since the fluctuations in the conditional distribution can be expected to average to zero over the ensemble so that

$$\frac{\partial p(\mathbf{x}, \dot{\mathbf{x}}, t)}{\partial t} = \int_{R_c} \left\{ \frac{\partial p(\mathbf{x}, \dot{\mathbf{x}}, t | \mathbf{c})}{\partial t} \right\} p(\mathbf{c}) d\mathbf{c} \rightarrow \mathbf{0}. \quad (58)$$

The solution to the stationary, unforced, and undamped form of Eq. (45) will depend upon the initial conditions. It can readily be verified that possible solutions are

$$p_s(\mathbf{x}, \dot{\mathbf{x}}) = A \int_{R_c} r \left[(1/2) \dot{x}_j \dot{x}_j + U(\mathbf{x}, \mathbf{c}) \right] p(\mathbf{c}) d\mathbf{c}, \quad (59a)$$

$$p_s(\mathbf{x}, \dot{\mathbf{x}}) = A \int_{R_c} \exp \left\{ -\alpha \left[(1/2) \dot{x}_j \dot{x}_j + U(\mathbf{x}, \mathbf{c}) \right] \right\} p(\mathbf{c}) d\mathbf{c}, \quad (59b)$$

where $U(\mathbf{x}, \mathbf{c})$ is the potential energy function, as in Eq. (37), and r is an arbitrary function. When r is the exponential function, as in Eq. (59b), the distribution of the system velocities has the same form as the Maxwell-Boltzmann distribution. More generally, for an integrable system it can be expected that the effect of averaging the conditional distribution $p(\mathbf{x}, \dot{\mathbf{x}}, t | \mathbf{c})$ across the ensemble to yield the unconditional distribution will be analogous to coarse graining, averaging out any variations in $\boldsymbol{\theta}$ -space and resulting in a jpdf that depends upon the action variables alone. This yields

$$p_s(\mathbf{J}, \boldsymbol{\theta}) = \int_{R_c} p(\mathbf{J}, \boldsymbol{\theta} | \mathbf{c}) p(\mathbf{c}) d\mathbf{c} = (2\pi)^{-N} p_s(\mathbf{J}). \quad (60)$$

Equations (59a,b) are in fact examples of Eq. (60), corresponding to the case in which the first action variable is equal to the energy of the system and the jpdf depends on this action variable alone. In general the final stationary distribution will depend on the initial distribution assigned to the action variables. The final entropy (of whatever type) will be

greater than the initial value, since the uniform distribution over θ represented by Eq. (60) is a maximum entropy condition. This feature is illustrated in Figure 7 for the 10 oscillator system that was considered in Section 3.2 and Figure 3. The Shannon and Boltzman-type entropies previously computed for the deterministic system are shown, together with the entropies associated with a random system in which the natural frequencies have the form $\omega_j = j + (0.5 - \theta_j)$ where θ_j is a random number that is uniformly distributed between 0 and 1. It is clear that the entropies of the random system tend to stationary values that exceed the initial values; the small fluctuations remaining in the figure arise from the use of a finite sample size in the simulations (30 samples).

In the case of a chaotic system, the discussion given at the end of section 3.2 can be expanded to consider the effects of system randomness. It is not possible to reach any general conclusions, although for a system that exhibits strong mixing it can be expected that the final stationary solution will be largely independent of the details of the starting distribution. If the system exhibits non-chaotic islands in phase space then the location of such islands will vary randomly across the ensemble of systems; in this case ensemble averaging may alleviate the long-term fluctuations associated with these islands and yield a stationary distribution within a relatively short timescale. Further consideration of this highly complex topic is beyond the scope of the present work.

It can be concluded that for a random system the entropy (of whatever type) can increase beyond an initial value, and the entropy can be expected to reach a stationary value at late times. Towards the end of section 3.2 it was stated that necessary conditions for an unforced and undamped deterministic system to achieve a stationary distribution are nonlinearity and coarse graining. From the preceding discussion it can be seen that neither condition is necessary in the case of a random system; for a deterministic system with random starting conditions the Zermelo (recurrence) paradox is overcome by nonlinearity and the Loschmidt (time reversal) paradox is overcome by coarse graining, but both paradoxes are overcome by system randomness regardless of nonlinearity or coarse graining.

4.3 A damped, forced system

When damping and forcing are present, Eq. (45) will yield a stationary solution that is independent of the starting conditions. It was shown in Section 3.3 that a Maxwell-Boltzmann distribution will be obtained for a deterministic system if the cross-spectral matrix

of the excitation forces is proportional to the damping matrix. In a similar way, it can readily be shown that if the conditional expectations of these items satisfy the condition

$$\bar{B}_{jk} = \left(\frac{2}{\alpha} \right) \bar{C}_{jk}, \quad (61)$$

then the solution to Eq. (45) will have exactly the form of Eq. (59b). The question then arises as to whether Eq. (61) can be expected to apply under fairly general conditions, or whether the equation represents a very special case. Consider initially the case of a linear system with an excitation applied at a single physical location. If the response is expressed in modal coordinates then the generalised force in mode j will be proportional to the modal displacement at the forcing point, ϕ_j say, and it then follows that $B_{jk}(\mathbf{c})$ will be proportional to $\phi_j \phi_k$. Given that the modal displacements can be either positive or negative, it is reasonable to assume that $\phi_j \phi_k$ will average to zero across the ensemble when $j \neq k$; furthermore, for $j = k$ the average value of ϕ_j^2 can be expected to be at least approximately independent of j , meaning that \bar{B}_{jk} will be proportional to the identity matrix. If this result is also approximately true of \bar{C}_{jk} (uniform, uncoupled damping) then Eq. (61) will apply and the response can be expected to have a Maxwell-Boltzmann distribution. It is well known that the higher frequency modes of a random system tend to have the eigen-statistics associated with the Gaussian Orthogonal Ensemble (GOE) family of random matrices [34]; in this case the eigenvectors (mode shapes) have independent and identically distributed entries, meaning that the conditions outlined above are fully satisfied, and Eq. (61) can be expected to apply to systems of this type. This argument can readily be extended to more general spatial loading conditions, and to nonlinear systems if the response is expressed in terms of a set of random global shape functions. It can therefore be stated as a general principle that system randomness will tend to promote the validity of Eq. (61) and hence the occurrence of the Maxwell-Boltzmann distribution. It is important to note that no claim is being made for the Maxwell-Boltzmann distribution for individual realisations of the random system; rather, the distribution applies across the ensemble of random system responses. This tendency is exploited in the following section, where coupled systems are considered.

As a final comment in this section, it can be noted that the foregoing arguments in favour of a Maxwell-Boltzmann distribution can be extended to random linear systems subjected to harmonic excitation. If the system is highly random, meaning that the natural frequencies vary to a high degree across the ensemble, then the response statistics will be

insensitive to the precise frequency of the harmonic excitation. The actual response can therefore be replaced by an average over excitations within a frequency band surrounding the original frequency; applying the excitations simultaneously in an uncorrelated way is equivalent to band limited white noise excitation, and the foregoing arguments can therefore be applied to the response distribution. Note that for this chain of reasoning to hold, the response statistics must be independent of the frequency of the excitation over a bandwidth that is sufficient to contain a reasonable number of modes.

5. Two Coupled Subsystems

5.1 *Forced response with damping*

This section is concerned with the dynamics of two subsystems that are connected through a lossless coupling. The degrees-of-freedom of the two systems are represented by \mathbf{x} and \mathbf{y} respectively, and the equations of motion of the coupled system are taken to have the form

$$\dot{\mathbf{z}} = \frac{d}{dt} \begin{pmatrix} \mathbf{x} \\ \dot{\mathbf{x}} \\ \mathbf{y} \\ \dot{\mathbf{y}} \end{pmatrix} = \begin{pmatrix} \dot{\mathbf{x}} \\ -\mathbf{g}_x(\mathbf{x}, \dot{\mathbf{x}}) - \mathbf{f}_{xy}(\mathbf{x}, \mathbf{y}) \\ \dot{\mathbf{y}} \\ -\mathbf{g}_y(\mathbf{y}, \dot{\mathbf{y}}) - \mathbf{f}_{yx}(\mathbf{x}, \mathbf{y}) \end{pmatrix} + \begin{pmatrix} \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{G}_x & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{G}_y \end{pmatrix} \begin{pmatrix} \mathbf{0} \\ \mathbf{w}_x(t) \\ \mathbf{0} \\ \mathbf{w}_y(t) \end{pmatrix}, \quad (62)$$

where \mathbf{z} is the complete state vector. Equation (62) is an obvious generalisation of Eq. (19), with the addition of the interaction forces $\mathbf{f}_{xy}(\mathbf{x}, \mathbf{y})$ and $\mathbf{f}_{yx}(\mathbf{x}, \mathbf{y})$ between the two subsystems.

The FPK equation associated with Eq. (62) has the form

$$\frac{\partial p(\mathbf{z}, t)}{\partial t} = L_x[p(\mathbf{z}, t)] + L_y[p(\mathbf{z}, t)] + \frac{\partial}{\partial \dot{\mathbf{x}}_j} [f_{xy,j}(\mathbf{x}, \mathbf{y}) p(\mathbf{z}, t)] + \frac{\partial}{\partial \dot{\mathbf{y}}_j} [f_{yx,j}(\mathbf{x}, \mathbf{y}) p(\mathbf{z}, t)], \quad (63)$$

where L_x and L_y are the differential operators associated with the individual subsystems, as in Eq. (21), and the last two terms on the right hand side of the equation arise from the coupling forces. Equation (63) relates to a deterministic system subjected to random forcing; if the system has random properties then the terms \mathbf{B} and \mathbf{g} that appear in the equation (via L_x and L_y), together with the coupling terms \mathbf{f} , must be replaced by conditional average values, as in Eqs. (46) and (47). For simplicity the following analysis is restricted to the case of a deterministic system, although the effect of system randomness is discussed towards the end of the present section.

The entropies of the two subsystems can be written as

$$S_x(t) = - \int_{R_x} p(\mathbf{x}, \dot{\mathbf{x}}, t) \ln[p(\mathbf{x}, \dot{\mathbf{x}}, t)] d\mathbf{x} d\dot{\mathbf{x}}, \quad (64)$$

$$S_y(t) = - \int_{R_y} p(\mathbf{y}, \dot{\mathbf{y}}, t) \ln[p(\mathbf{y}, \dot{\mathbf{y}}, t)] d\mathbf{y} d\dot{\mathbf{y}}, \quad (65)$$

and the entropy of the complete system has the form

$$S_z(t) = - \int_{R_z} p(\mathbf{z}, t) \ln[p(\mathbf{z}, t)] d\mathbf{z}. \quad (66)$$

It follows from Eq. (64) that

$$\dot{S}_x(t) = - \int_{R_z} \{ \ln[p(\mathbf{x}, \dot{\mathbf{x}}, t)] + 1 \} \frac{\partial p(\mathbf{z}, t)}{\partial t} d\mathbf{z}, \quad (67)$$

where the normal integration over \mathbf{x} and $\dot{\mathbf{x}}$, as in Eq. (23), has been replaced by integration over \mathbf{z} to aid the following algebraic development. Specifically, it follows from Eqs. (63) and (67) that

$$\dot{S}_x(t) = -C_{x,jj} + F_x(t) + \int_{R_z} \frac{\partial \ln p(\mathbf{x}, \dot{\mathbf{x}}, t)}{\partial \dot{x}_j} f_{xy,j}(\mathbf{x}, \mathbf{y}) p(\mathbf{z}, t) d\mathbf{z}, \quad (68)$$

$$F_x(t) = \frac{1}{2} \int_{R_x} \left\{ \frac{B_{x,jk}}{p(\mathbf{x}, \dot{\mathbf{x}}, t)} \frac{\partial p(\mathbf{x}, \dot{\mathbf{x}}, t)}{\partial \dot{x}_j} \frac{\partial p(\mathbf{x}, \dot{\mathbf{x}}, t)}{\partial \dot{x}_k} \right\} d\mathbf{x} d\dot{\mathbf{x}}. \quad (69)$$

By analogy, the rate of change of the entropy of the second subsystem can be written as

$$\dot{S}_y(t) = -C_{y,jj} + F_y(t) + \int_{R_z} \frac{\partial \ln p(\mathbf{y}, \dot{\mathbf{y}}, t)}{\partial \dot{y}_j} f_{yx,j}(\mathbf{x}, \mathbf{y}) p(\mathbf{z}, t) d\mathbf{z}, \quad (70)$$

$$F_y(t) = \frac{1}{2} \int_{R_y} \left\{ \frac{B_{y,jk}}{p(\mathbf{y}, \dot{\mathbf{y}}, t)} \frac{\partial p(\mathbf{y}, \dot{\mathbf{y}}, t)}{\partial \dot{y}_j} \frac{\partial p(\mathbf{y}, \dot{\mathbf{y}}, t)}{\partial \dot{y}_k} \right\} d\mathbf{y} d\dot{\mathbf{y}}. \quad (71)$$

In addition, the rate of change of the entropy of the complete system can be found from Eq. (66) to have the form

$$\dot{S}_z(t) = -C_{x,jj} - C_{y,jj} + F_z(t), \quad (72)$$

$$F_z(t) = \frac{1}{2} \int_{R_z} \left\{ \frac{B_{x,jk}}{p(\mathbf{z}, t)} \frac{\partial p(\mathbf{z}, t)}{\partial \dot{x}_j} \frac{\partial p(\mathbf{z}, t)}{\partial \dot{x}_k} \right\} d\mathbf{z} + \frac{1}{2} \int_{R_z} \left\{ \frac{B_{y,jk}}{p(\mathbf{z}, t)} \frac{\partial p(\mathbf{z}, t)}{\partial \dot{y}_j} \frac{\partial p(\mathbf{z}, t)}{\partial \dot{y}_k} \right\} d\mathbf{z}. \quad (73)$$

It is shown in the Appendix that the Fisher information has the property

$$F_z \geq F_x + F_y, \quad (74)$$

and in combination with Eqs. (68), (70), and (72) this result can be used to study the statistical mechanics of the late time *stationary case*, in which the rate of change of all entropy measures is zero. In particular, it follows from the equations that

$$\int_{R_z} \frac{\partial \ln p(\mathbf{x}, \dot{\mathbf{x}}, t)}{\partial \dot{x}_j} f_{xy,j}(\mathbf{x}, \mathbf{y}) p(\mathbf{z}, t) d\mathbf{z} + \int_{R_z} \frac{\partial \ln p(\mathbf{y}, \dot{\mathbf{y}}, t)}{\partial \dot{y}_j} f_{yx,j}(\mathbf{x}, \mathbf{y}) p(\mathbf{z}, t) d\mathbf{z} \geq 0. \quad (75)$$

This result does not have an immediate physical interpretation, but progress can be made by considering the case in which the response in each subsystem has a Maxwell-Boltzmann distribution, Eq. (38), so that for example

$$\frac{\partial \ln p(\mathbf{x}, \dot{\mathbf{x}}, t)}{\partial \dot{x}_j} = -\frac{\dot{x}_j}{\sigma_x^2}, \quad (76)$$

where $\sigma_x^2 = 1/\alpha$ is the mean squared velocity of each degree of freedom in the first subsystem. Equations (75) and (76) then yield

$$dq \left(\frac{1}{\sigma_y^2} - \frac{1}{\sigma_x^2} \right) \geq 0, \quad (77)$$

$$dq = E[\dot{x}_j f_{xy,j}(\mathbf{x}, \mathbf{y})] = -E[\dot{y}_j f_{yx,j}(\mathbf{x}, \mathbf{y})]. \quad (78)$$

The term dq is the ensemble average value of the power leaving the x -subsystem through the coupling (the sum of the coupling forces times the relevant velocities); since the coupling is conservative, this is also equal to the power received by the y -subsystem through the coupling, as indicated by the final expression in Eq. (78). It can be noted that dq is *not* an incremental quantity, but this notation has been adopted here to make a connection with the notation for heat exchange employed in classical thermodynamics. Although Eqs. (77) and (78) have been derived on the assumption that the distribution of the velocities in each subsystem is of Maxwell-type, the equations may still be valid if this assumption is relaxed to some extent. The summation over the index j that appears in Eqs. (75) and (78) renders the equation robust against deviations from the Maxwell-Boltzmann distribution – i.e. the distribution only has to apply in an average sense, and σ_x^2 (for example) can be interpreted as the “average” mean squared velocity of the subsystem rather than the mean squared velocity of each individual degree of freedom. The assumption of a Maxwell-Boltzmann distribution (or near approximation) can be justified for a deterministic system on the basis of section 3.3, providing Eq. (39) holds and the coupling between the two subsystems is weak so that the response distribution is not strongly affected by the coupling. For a random system the justification is more robust, since in this case the condition given by Eq. (39) is replaced by the ensemble average condition represented by Eq. (61); as discussed in section 4.3 this condition is likely to be met in a wide range of cases.

If σ_x^2 and σ_y^2 are interpreted as measures of the “temperature” of each subsystem, then Eq. (77) implies that energy will flow from the “hotter” subsystem to the “cooler” subsystem. It is notable that this result has been derived here for a system that is not closed

(the system has both forcing and damping) and that the inequality in Eq. (77) has resulted not from a property of entropy but from a property of Fisher information, Eq. (74). From a vibrational point of view the “temperature” σ_x^2 is actually related to the kinetic energy per degree of freedom, which can be interpreted spatially as the kinetic energy density, or for a linear system as the kinetic energy per mode. In deriving Eq. (77) no assumption has been made regarding the linearity of the system, but in the special case of a linear system further consequences of the equation can be explored. For a linear system the energy flow is a second order function of the response, and the ensemble average energy flow will therefore be a linear function of the response covariance matrix. For the Maxwell-Boltzmann distribution the correlation between displacement and velocity is zero, and moreover for a linear system the mean squared displacements can be linearly expressed in terms of the mean squared velocities. This implies that the energy flow is a linear function of σ_x^2 and σ_y^2 . The only linear function that will ensure the validity of Eq. (77) for all values of σ_x^2 and σ_y^2 is

$$dq = \beta(\sigma_x^2 - \sigma_y^2), \quad (79)$$

for some constant β . Equation (79), which follows from Eq. (77) for a linear system by the foregoing argument, is actually a statement of the fundamental tenet of the Statistical Energy Analysis (SEA) approach to high frequency vibration analysis: the energy flow between two coupled subsystems is proportional to the difference in the modal energy of the two subsystems (i.e. the energy per mode). A power balance for each subsystem then yields the full set of SEA equations in the standard form [12]

$$\omega\eta_x(N_x\sigma_x^2) + \beta(\sigma_x^2 - \sigma_y^2) = P_x, \quad (80)$$

$$\omega\eta_y(N_y\sigma_y^2) + \beta(\sigma_y^2 - \sigma_x^2) = P_y, \quad (81)$$

where, for example, N_x is the number of degrees of freedom in the first subsystem and P_x is the ensemble average of the power input arising from the applied loading. The first term on the left hand side of Eq. (80) represents the ensemble average of the power dissipated by the subsystem, where ω is a nominal centre frequency of the excitation, η_x is the effective loss factor, and $N_x\sigma_x^2$ is the ensemble average of the energy of the subsystem.

There has been much discussion in the literature concerning the conditions which are required for the SEA equations to be valid (see for example reference [35]). If the degrees of freedom are expressed in terms of the uncoupled modes of each subsystem, then it is often stated that the modal responses in any one subsystem should be uncorrelated and should have

equipartition of energy; furthermore, there should be no correlation of the responses in different subsystems. For continuous subsystems these conditions can be restated in terms of wave motion: each subsystem should carry a diffuse wavefield, and the wavefields should be uncorrelated across different subsystems. These conditions (whether expressed in terms of modes or waves) tend to be met if the coupling between the subsystems is weak, and if system randomness is present, so that correlations might average to zero across the ensemble of systems (see for example references [14,36]). The present derivation of Eq. (79) is fully consistent with these findings, in that: (i) the Maxwell-Boltzmann distribution has equipartition of kinetic energy among the subsystem degrees of freedom, which are uncorrelated, (ii) the Maxwell-Boltzmann distribution is promoted by system randomness, (ii) the Maxwell-Boltzmann distribution is promoted by weak coupling, in the sense that the distribution of the response will be close to that of the uncoupled subsystem. The present approach therefore offers an alternative route to deriving both the SEA equations and the conditions for their validity.

The foregoing discussion regarding SEA relates to a linear system, but it can also be noted that in deriving Eq. (77), as opposed to Eq. (79), no assumption has been made regarding the linearity of the system. The fact that the “temperature” of a subsystem is proportional to the average kinetic energy (rather than the potential energy or the total energy) is consistent with the tenets of statistical thermodynamics [1] and ab initio molecular simulations [37], and arises from applying the Maxwell-Boltzmann distribution to the classical thermodynamic relation $TdS = dE$. In detail, if the entropy associated with the Maxwell-Boltzmann distribution, Eq. (38), is differentiated with respect to the *total* average energy then the result is α , and α is equal to $1/\sigma_x^2$, meaning that $T = \sigma_x^2$. This is in agreement with Eq. (77) and implies that the temperature is given by the average kinetic energy, regardless of any system nonlinearity. In contrast, Carcaterra [38] has expressed the temperature of a nonlinear vibrating system as a nonlinear function of the total energy; such an approach is not inconsistent with the present analysis, in that the kinetic energy can be considered to be a function of the total energy. For a linear system the average kinetic energy will be equal to half the total energy, but for a nonlinear system a nonlinear relationship can arise, and this is the case for the systems considered in reference [38].

5.2 An undamped, unforced system

In the absence of damping and forcing, Eqs. (68), (70), and (72) become respectively

$$\dot{S}_x(t) = \int_{R_z} \frac{\partial \ln p(\mathbf{x}, \dot{\mathbf{x}}, t)}{\partial \dot{x}_j} f_{xy,j}(\mathbf{x}, \mathbf{y}) p(\mathbf{z}, t) d\mathbf{z}, \quad (82)$$

$$\dot{S}_y(t) = \int_{R_z} \frac{\partial \ln p(\mathbf{y}, \dot{\mathbf{y}}, t)}{\partial \dot{y}_j} f_{yx,j}(\mathbf{x}, \mathbf{y}) p(\mathbf{z}, t) d\mathbf{z}, \quad (83)$$

$$\dot{S}_z(t) = 0. \quad (84)$$

Given that entropy is always reduced by correlation between variables it follows that

$$S_z(t) \leq S_x(t) + S_y(t), \quad (85)$$

and if the two subsystems are uncorrelated at $t = 0$ then

$$S_z(0) = S_x(0) + S_y(0). \quad (86)$$

Equations (83) and (84) imply that

$$\int_0^t \{ \dot{S}_x(t) + \dot{S}_y(t) \} dt \geq 0. \quad (87)$$

To explore the potential implications of Eq. (87), initially an idealized case will be considered in which each of the two subsystems is taken to have a near-Maxwell distribution of response at all times. Equations (82), (83) and (87) then yield

$$\int_0^t \left\{ dq \left(\frac{1}{\sigma_y^2} - \frac{1}{\sigma_x^2} \right) \right\} dt \geq 0, \quad (88)$$

where dq is given by Eq. (78) and the subsystem “temperatures” σ_x^2 and σ_y^2 will vary with time. If the number of degrees of freedom in the two subsystems is respectively N_x and N_y then the ensemble average of the kinetic energy of the whole system is given by

$$T = (N_x \sigma_x^2 + N_y \sigma_y^2) / 2. \quad (89)$$

Now if the system is additionally assumed to be linear, then the total energy will be equal to twice the kinetic energy, and conservation of energy then implies that

$$dq = -N_x \frac{d\sigma_x^2}{dt} = N_y \frac{d\sigma_y^2}{dt}. \quad (90)$$

If Eq. (90) is substituted into Eq. (88) then the integral can be performed to yield

$$\sigma_x^{2N_x}(t) \sigma_y^{2N_y}(t) \geq \sigma_x^{2N_x}(0) \sigma_y^{2N_y}(0). \quad (91)$$

This result does not have an immediate physical interpretation, but an insight can be obtained by considering the conditions under which the quantity $\sigma_x^{2N_x} \sigma_y^{2N_y}$ is maximised subject to the

system having a prescribed value of energy. This can be explored by maximising the function

$$V = \sigma_x^{2N_x} \sigma_y^{2N_y} - \lambda(N_x \sigma_x^2 + N_y \sigma_y^2), \quad (92)$$

where λ is a Lagrange multiplier. In this way it is readily found that the maximum value of $\sigma_x^{2N_x} \sigma_y^{2N_y}$ is obtained when $\sigma_x^2 = \sigma_y^2$, i.e. when the “temperatures” of the two systems are the same. Equation (91) can therefore be interpreted as a statement that the two subsystems are nearer to a state of equal temperature at time t than at the initial time. However, this *does not* imply that the system will reach a steady-state condition, or that the temperatures will be equal should such a condition be reached. If an additional assumption is made that the energy flow between the subsystems is a linear function of σ_x^2 and σ_y^2 , then the fact that Eq. (88) holds for small times t and for any starting values of σ_x^2 and σ_y^2 leads to the conclusion that $dq = \beta(\sigma_x^2 - \sigma_y^2)$, as in Eq. (79). Equation (92) then gives

$$-N_x \frac{d\sigma_x^2}{dt} = N_y \frac{d\sigma_y^2}{dt} = \beta(\sigma_x^2 - \sigma_y^2). \quad (93)$$

These equations are in fact identical to the Transient Statistical Energy Analysis (TSEA) equations that are employed in high frequency vibration analysis, and they imply (subject to the underlying assumptions) that the system *does* reach a steady state condition with $\sigma_x^2 = \sigma_y^2$.

To summarise the foregoing discussion, it can be deduced from entropy considerations that Eq. (88) must apply if the two subsystems have a near-Maxwell distribution. As in the previous section, “near-Maxwell” can be interpreted as a requirement that Eq. (76) is a good approximation on average across the various degrees of freedom in a subsystem. If the system is linear then Eq. (88) implies that the subsystems will be closer to having equality of temperature at any time $t > 0$ than at the initial state, at least in the sense of Eq. (91). Furthermore, if the energy flow is assumed to be a linear function of the temperatures then TSEA is recovered, and the system tends to a steady-state condition in which the temperatures of the two subsystems are equal. System variability and ensemble averaging can be expected to promote the occurrence of a near-Maxwell distribution, meaning that these results are more likely to apply to random systems. In addition, the transition to a stationary distribution (but not necessarily an equal temperature distribution) can be expected for a random system on the basis of Eq. (58).

In contrast to the above comments, it is known that there are cases in which coupled subsystems *no not* reach an equal temperature state, even if the system is linear [39,40]. This

tends to occur when there is strong coupling between the subsystems, and there is a statistical correlation between the two subsystems which can be related to the degree of mode localisation [39]. This strong correlation violates the assumption of a near-Maxwell distribution in each subsystem, and also the assumption that the energy flow can be expressed solely as a linear function of σ_x^2 and σ_y^2 .

6. Conclusions

The main results of the present work can be summarised by considering the different types of system that have been considered. These are listed below, together with the pertinent findings.

1) A closed deterministic system. This type of system was considered in Section 3.2, where linear, integrable, and chaotic systems were each discussed. In the case of a *linear* system the angular rates $\dot{\boldsymbol{\theta}} = \boldsymbol{\omega}(\mathbf{J})$ are the same for all values of the action variables \mathbf{J} and the system is uniformly recurrent. This means that all possible measures of entropy will recur, and the system cannot achieve a stationary state. The recurrence time may be extremely long but the key point is that the system will at some point return to the initial distribution. For a *nonlinear integrable* system the situation is different, in that the angular rates depend on the action variables, and the system is not uniformly recurrent, meaning that an initial distribution will not in general be recovered. For a given set of action variables \mathbf{J} the trajectories of the system in the angle space $\boldsymbol{\theta}$ are non-dispersive, and hence the response distribution cannot achieve a strictly stationary state. However, the coarse grained jpdf of the system will become uniform across the angle space at large times, and hence a stationary coarse grained distribution can be obtained; this was illustrated for a simple duffing oscillator in Section 3.2. The stationary distribution will depend on the initial distribution assigned to the action variables. For a *chaotic* system the initial distribution will spread and become filamentary across phase space, and in many cases it can be expected that coarse graining will lead to a stationary distribution which will depend on the initial distribution of the system energy. The situation can be much more complicated however, as discussed in reference [5]. The stationary coarse grained entropy of an integrable or chaotic system will be greater than the initial value, although there is no reason why the increase in entropy should be monotonic. In summary, for a deterministic system nonlinearity and coarse graining are both necessary for a stationary value of the entropy to be achieved.

2) A closed random system. This case was considered in Section 4.2, where it was shown that system randomness can lead to a stationary jpdf of the response, and hence a stationary value of the entropy, even for a linear system. This result applies to all the considered definitions of the entropy, including the Shannon entropy. System randomness (or equivalently, randomness in the equations of motion) is a very natural concept in engineering, where manufacturing imperfections and variable material properties can be expected for complex built-up systems. This is not the case in classical statistical mechanics, where the physical systems under consideration are taken to have well defined Hamiltonians. The stationary late-time value of the entropy resulting from system randomness will be greater than the initial value, but the increase in entropy is unlikely to be monotonic, and the final stationary distribution will depend on the initial distribution, as indicated in Eq. (60) for the case of an integrable system.

3) A forced-damped deterministic system. It was shown in Section 3.3 that the stationary response of an open system will have a Maxwell-Boltzmann distribution if the forcing and damping acting on the system are related via Eq. (39): the cross-spectral matrix of the excitation must be proportional to the damping matrix.

4) A forced-damped random system. In this case the stationary response will have a Maxwell-Boltzmann distribution if Eq. (39) is satisfied across the ensemble of systems in the average sense of Eq. (61). This is a much more relaxed condition than Eq. (39), and it is argued in Section 4.3 that (for example) the condition is likely to be met for a linear system with highly random mode shapes. So again it can be seen that the system randomness that can occur in engineering can produce behaviour which would not be expected so readily in statistical mechanics.

5) Closed coupled systems. It is shown in section 5.2 that if two coupled subsystems each have a near-Maxwell distribution then the energy flow between the systems tends to promote equal temperatures, at least in the generalised sense of Eq. (90). The “temperature” here is defined in terms of the average kinetic energy of a system, as in statistical mechanics, and not in terms of the total energy as considered in previous work on statistical structural dynamics [38]. If the two subsystems are linear, and the associated Maxwell-Boltzmann distributions are assumed to have very weak correlation between the subsystems, then the equations lead

to an emergent law that is identical to Transient Statistical Energy Analysis (TSEA). It can be noted from point (4) that the assumption of the occurrence of near-Maxwell distributions is more likely to be justified for random systems.

6) Forced-damped coupled systems. It is shown in section (5.1) that in the stationary state the energy flow between two coupled systems is from the higher temperature subsystem to the lower temperature subsystem. Although this result is hardly surprising, it should be noted that the result has been derived here from the fundamental equations of motion of the system, with no recourse to the concepts of classical thermodynamics. If the subsystems are linear then the emergent law for the flow of energy is identical to Statistical Energy Analysis (SEA). As in point (5), these results are based on the occurrence of near-Maxwell distributions, and this can be justified for random systems on the basis of point (4). It is curious to note that the energy flow law, Eq. (77), emerges from the properties of Fisher information rather than the properties of entropy, although the present theoretical framework arises from applying the condition of constant entropy in the stationary state.

It can be noted that the conclusions listed in point (1) are not new to the statistical mechanics community. The remaining conclusions are based to a greater or lesser degree on system randomness, and this represents a departure from the normal approach taken in statistical mechanics (although, as an aside, it can be argued that the well know Kac model [41] in statistical mechanics contains an aspect that can be interpreted either as system randomness or random initial conditions, namely the placement of certain “colour changing” barriers that act on a set of balls that circulate in discrete steps; however this model still exhibits recurrence, whereas a truly random system does not). In summary the work has shown how the properties of entropy and Fisher information can be used in conjunction with the fundamental equations of motion of a structural dynamic system to derive key properties of the system, ranging from stationary distributions to emergent laws. As an example of the value of this approach, the energy flow laws represented by Eqs. (77) and (78) are not limited to linear systems, and these equations could not be derived from any of the standard approaches used to justify the SEA equations. The approach therefore provides a methodology for the systematic study of energy flow in nonlinear systems.

Appendix: Properties of the Fisher Information

Consider a coupled system whose state vector can be written in the form

$$\mathbf{z} = \begin{pmatrix} \hat{\mathbf{x}} \\ \hat{\mathbf{y}} \end{pmatrix}, \quad (\text{A1})$$

where $\hat{\mathbf{x}}$ and $\hat{\mathbf{y}}$ are the state vectors of the two constituent subsystems. The Fisher information for each subsystem is defined by omitting correlations between $\hat{\mathbf{x}}$ and $\hat{\mathbf{y}}$ so that

$$F_x(t) = \frac{1}{2} \int_{R_x} \frac{1}{p(\hat{\mathbf{x}}, t)} \frac{\partial p(\hat{\mathbf{x}}, t)}{\partial \dot{x}_j} \frac{\partial}{\partial \dot{x}_k} \{B_{x,jk}(\hat{\mathbf{x}}) p(\hat{\mathbf{x}}, t)\} d\hat{\mathbf{x}}, \quad (\text{A2})$$

$$F_y(t) = \frac{1}{2} \int_{R_y} \frac{1}{p(\hat{\mathbf{y}}, t)} \frac{\partial p(\hat{\mathbf{y}}, t)}{\partial \dot{y}_j} \frac{\partial}{\partial \dot{y}_k} \{B_{y,jk}(\hat{\mathbf{y}}) p(\hat{\mathbf{y}}, t)\} d\hat{\mathbf{y}}. \quad (\text{A3})$$

Conversely, the Fisher information for the whole system includes the effect of correlations, so that

$$F_z(t) = \frac{1}{2} \int_{R_z} \frac{1}{p(\mathbf{z}, t)} \frac{\partial p(\mathbf{z}, t)}{\partial \dot{x}_j} \frac{\partial}{\partial \dot{x}_k} \{B_{x,jk}(\hat{\mathbf{x}}) p(\mathbf{z}, t)\} d\mathbf{z} + \frac{1}{2} \int_{R_z} \frac{1}{p(\mathbf{z}, t)} \frac{\partial p(\mathbf{z}, t)}{\partial \dot{y}_j} \frac{\partial}{\partial \dot{y}_k} \{B_{y,jk}(\hat{\mathbf{y}}) p(\mathbf{z}, t)\} d\mathbf{z} \quad (\text{A4})$$

Now Eq. (A4) can be written in the form

$$F_z(t) = I_x + I_y, \quad (\text{A5})$$

where (for example)

$$I_x = \frac{1}{2} \int_{R_z} \frac{1}{p(\hat{\mathbf{x}}, \hat{\mathbf{y}})} \frac{\partial p(\hat{\mathbf{x}}, \hat{\mathbf{y}})}{\partial \dot{x}_j} \frac{\partial}{\partial \dot{x}_k} \{B_{x,jk}(\hat{\mathbf{x}}) p(\hat{\mathbf{x}}, \hat{\mathbf{y}})\} d\mathbf{z}. \quad (\text{A6})$$

For ease of notation, explicit dependencies on t are omitted from the above equation and in the equations that follow. Noting that

$$p(\hat{\mathbf{x}}, \hat{\mathbf{y}}) = p(\hat{\mathbf{x}}) p(\hat{\mathbf{y}} | \hat{\mathbf{x}}), \quad (\text{A7})$$

it follows that

$$\frac{\partial p(\hat{\mathbf{x}}, \hat{\mathbf{y}})}{\partial \dot{x}_j} = \frac{\partial p(\hat{\mathbf{x}})}{\partial \dot{x}_j} p(\hat{\mathbf{y}} | \hat{\mathbf{x}}) + p(\hat{\mathbf{x}}) \frac{\partial p(\hat{\mathbf{y}} | \hat{\mathbf{x}})}{\partial \dot{x}_j}, \quad (\text{A8})$$

$$\frac{\partial}{\partial \dot{x}_k} \{B_{x,jk}(\hat{\mathbf{x}}) p(\hat{\mathbf{x}}, \hat{\mathbf{y}})\} = p(\hat{\mathbf{y}} | \hat{\mathbf{x}}) \frac{\partial}{\partial \dot{x}_k} \{B_{x,jk}(\hat{\mathbf{x}}) p(\hat{\mathbf{x}})\} + B_{x,jk}(\hat{\mathbf{x}}) p(\hat{\mathbf{x}}) \frac{\partial p(\hat{\mathbf{y}} | \hat{\mathbf{x}})}{\partial \dot{x}_k}. \quad (\text{A9})$$

Equation (A6) can now be written as the sum of four terms

$$I_x = I_{x1} + I_{x2} + I_{x3} + I_{x4}, \quad (\text{A10})$$

where

$$I_{x1} = \frac{1}{2} \int_{R_z} \frac{p(\hat{\mathbf{y}}|\hat{\mathbf{x}})}{p(\hat{\mathbf{x}})} \frac{\partial p(\hat{\mathbf{x}})}{\partial \dot{x}_j} \frac{\partial}{\partial \dot{x}_k} \{B_{x,jk}(\hat{\mathbf{x}}) p(\hat{\mathbf{x}})\} d\mathbf{z} = F_x, \quad (\text{A11})$$

$$I_{x2} = \frac{1}{2} \int_{R_z} B_{x,jk}(\hat{\mathbf{x}}) \left\{ \frac{\partial p(\hat{\mathbf{x}})}{\partial \dot{x}_j} \frac{\partial p(\hat{\mathbf{y}}|\hat{\mathbf{x}})}{\partial \dot{x}_k} \right\} d\mathbf{z} = 0, \quad (\text{A12})$$

$$I_{x3} = \frac{1}{2} \int_{R_z} \frac{\partial p(\hat{\mathbf{y}}|\hat{\mathbf{x}})}{\partial \dot{x}_j} \frac{\partial}{\partial \dot{x}_k} \{B_{x,jk}(\hat{\mathbf{x}}) p(\hat{\mathbf{x}})\} d\mathbf{z} = 0, \quad (\text{A13})$$

$$I_{x4} = \frac{1}{2} \int_{R_z} B_{x,jk}(\hat{\mathbf{x}}) \left\{ \frac{p(\hat{\mathbf{x}})}{p(\hat{\mathbf{y}}|\hat{\mathbf{x}})} \frac{\partial p(\hat{\mathbf{y}}|\hat{\mathbf{x}})}{\partial \dot{x}_j} \frac{\partial p(\hat{\mathbf{y}}|\hat{\mathbf{x}})}{\partial \dot{x}_k} \right\} d\mathbf{z} \geq 0. \quad (\text{A14})$$

It has been noted in Eq. (A14) that \mathbf{B}_x is a positive semi-definite matrix. It then follows that

$$I_x \geq F_x. \quad (\text{A15})$$

An analogous result holds for I_y , and hence it follows that

$$F_z \geq F_x + F_y, \quad (\text{A16})$$

which means that any correlation between the subsystems leads to an increase in the Fisher information. This result is employed in Section 5.1.

References

- [1] F. Schwabl, Statistical Mechanics, Second Edition, Springer, Berlin, 2006.
- [2] M. Le Bellac, F. Mortessagne, G.G. Batrouni, Equilibrium and Non-Equilibrium Statistical Thermodynamics, Cambridge University Press, Cambridge, 2004.
- [3] J. Bricmont, Science of chaos or chaos in science? Annals of the New York Academy of Science, 775 (1996) 131-175.
- [4] N. Singh, On the foundations of statistical mechanics: a brief review, Modern Physics Letters B 27 (2013) 1330003/1-130003/29.
- [5] G.M. Zaslavsky, Chaotic dynamics and statistical laws, Physics Today 52 (1999) 39-45.
- [6] E.T. Jaynes, Probability Theory: the Logic of Science, Cambridge University Press, Cambridge, 2003.
- [7] J.W. Gibbs, The Collected Works, Volume II: Elementary Principles in Statistical Mechanics, Yale University Press, New Haven, 1948.
- [8] O. Darrigol, Atoms, Mechanics, and Probability: Ludwig Boltzmann's Statistico-Mechanical Writings – An Exegesis, Oxford University Press, Oxford, 2018.

- [9] P. Ehrenfest, T. Ehrenfest, *The Conceptual Foundations of the Statistical Approach in Mechanics*, Dover Publications Inc, Mineola, 1990.
- [10] L.D. Landau, E.M. Lifshitz, *Statistical Physics, Part 1*, Third Edition, Pergamon Press, Oxford, 1980.
- [11] A.J. Keane, P.B. Nair, *Computational Approaches for Aerospace Design (Chapter 8: Design in the Presence of Uncertainty)*, John Wiley and Sons, Chichester, 2005.
- [12] R.H. Lyon, R.G. DeJong, *Theory and Application of Statistical Energy Analysis*, Second Edition, Butterworth-Heinemann, Boston, 1995.
- [13] J. Woodhouse, An approach to the theoretical background to Statistical Energy Analysis applied to structural vibration. *J. Acoustical Society of America* 69 (1981) 1695-1709.
- [14] B.R. Mace, The statistics of power flow between two continuous one-dimensional subsystems, *J. Sound Vib.* 154 (1992) 321-341.
- [15] P.J. Shorter, R.S. Langley, Vibro-acoustic analysis of complex systems. *J. Sound Vib.* 288 (2005) 669-699.
- [16] A. Le Bot, Entropy in statistical energy analysis. *J. Acoustical Society of America* 125 (2009) 1473-1478.
- [17] A. Le Bot, *Foundations of Statistical Energy Analysis in Vibroacoustics*, Oxford University Press, Oxford, 2015.
- [18] W.M. Haddad, V. Chellaboina, S.G. Nersesov, *Thermodynamics: a Dynamical Systems Approach*, Princeton University Press, Princeton, 2005.
- [19] A. Carcaterra, An entropy formulation for the analysis of energy flow between mechanical resonators, *Mechanical Systems and Signal Processing* 16 (2002) 905-920.
- [20] A.I. Khinchin, *Mathematical Foundations of Statistical Mechanics*, Dover Publication, New York, 1949.
- [21] B.R. Frieden, *Science from Fisher Information*, Cambridge University Press, Cambridge, 2004.
- [22] R. Frigg, C. Werndl, Entropy - a guide for the perplexed, in C. Beisbart, S. Hartmann (Eds.), *Probabilities in Physics*, Oxford University Press, Oxford, 2011, 115-142.
- [23] D. Gillies, *Philosophical Theories of Probability*, Routledge, London, 2000.
- [24] R.B. Ash, *Information Theory*, Dover Publications Inc, New York, 1990.
- [25] Y.K. Lin, *Probabilistic Theory of Structural Dynamics*, McGraw-Hill, New York, 1967.

- [26] L. Wasserman, All of Statistics: a Concise Course in Statistical Inference, Springer, New York, 2004.
- [27] R.S. Langley, Probability functionals for self-consistent and invariant inference: entropy and fisher information, IEEE Transactions on Information Theory, 59 (2013) 4397-4407.
- [28] C.W. Gardiner, Handbook of Stochastic Methods, Second Edition, Springer-Verlag, Berlin, 1985.
- [29] H.J. Korsch, H. J. Jodl, T. Hartmann, Chaos (Chapter 8: The Duffing Oscillator), Springer, Berlin, 2008.
- [30] J.M. Zhang, Y. Liu, Witnessing a Poincare recurrence with Mathematica, Results in Physics 7 (2017) 3373-3379.
- [31] D. Wallace, Recurrence Theorems: a unified account, Journal of Mathematical Physics 56 (2015) 022105/1-022105/11.
- [32] T. Tel, M. Gruiz, Chaotic Dynamics, Cambridge University Press, Cambridge, 2006.
- [33] M. Viana, K. Oliveira, Foundations of Ergodic Theory, Cambridge University Press, Cambridge, 2016.
- [34] R.L. Weaver, The unreasonable effectiveness of random matrix theory for the vibrations and acoustics of complex structures, in: M.C.M. Wright, R. Weaver (Eds.), New Directions in Linear Acoustics and Vibration: Quantum Chaos, Random Matrix Theory, and Complexity, Cambridge University Press, Cambridge, 2010.
- [35] F.J. Fahy, Statistical energy analysis: a critical overview, Philosophical Transactions of the Royal Society: Physical Sciences and Engineering, 346 (1994) 431-447.
- [36] S. Finnveden, A quantitative criterion validating coupling power proportionality in statistical energy analysis, J. Sound Vib. 330 (2011) 87-109
- [37] D. Frenkel, B. Smit, Understanding Molecular Simulation, Academic Press, San Diego, 2002.
- [38] A. Carcaterra, Thermodynamic temperature in linear and nonlinear Hamiltonian systems, International Journal of Engineering Science, 80 (2014) 189-208.
- [39] B.R. Mace, Statistical energy analysis: Coupling loss factors, indirect coupling and system modes, J. Sound Vib. 279 (2005) 141-170.
- [40] R.L. Weaver, O.I. Lobkis, Anderson Localization in Coupled Reverberation Rooms, J. Sound Vib. 231 (2000) 1111-1134.
- [41] M. Kac, Probability and Related Topics in Physical Sciences, Interscience, New York, 1959.

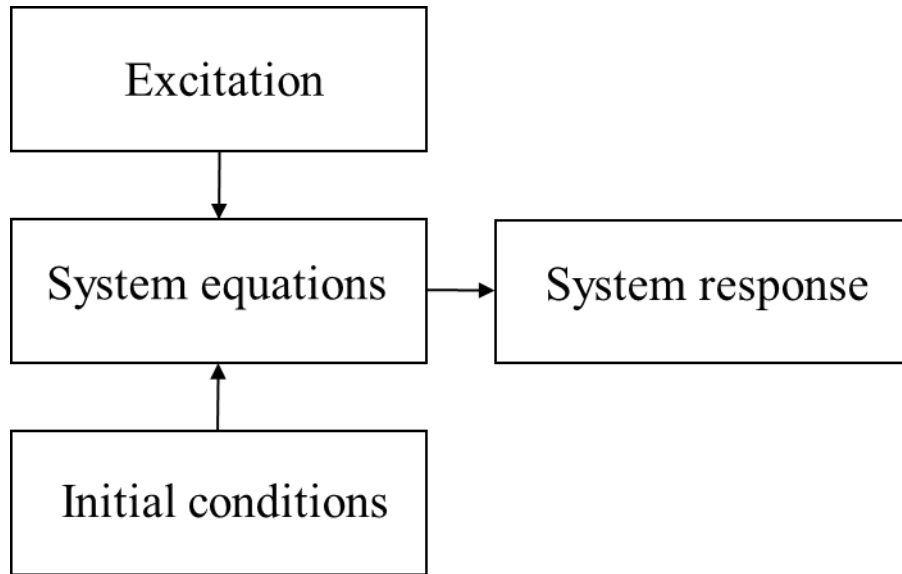


Figure 1 The components of an engineering vibration analysis

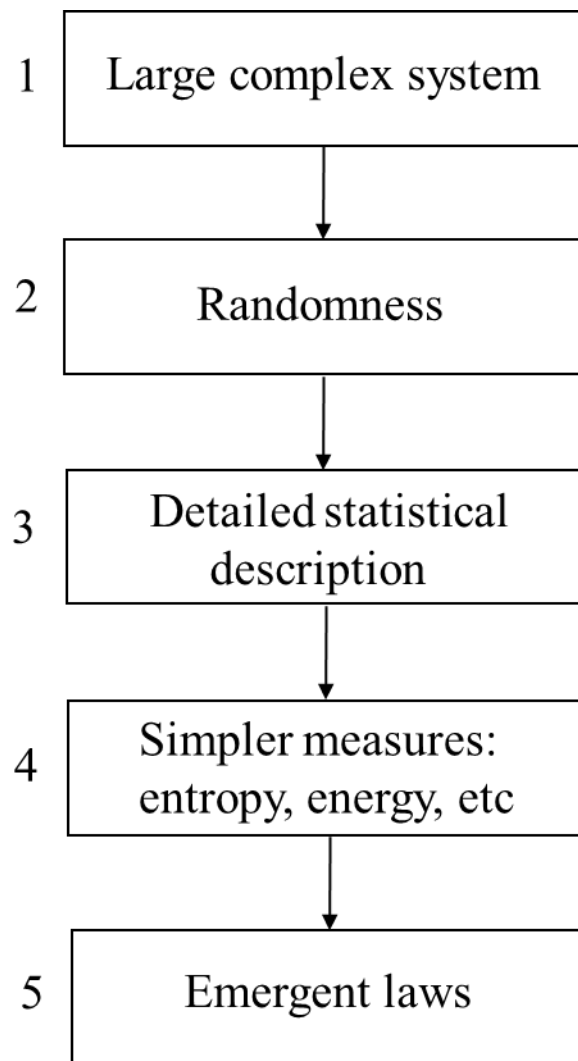


Figure 2 A simplified outline of the research agenda of statistical mechanics

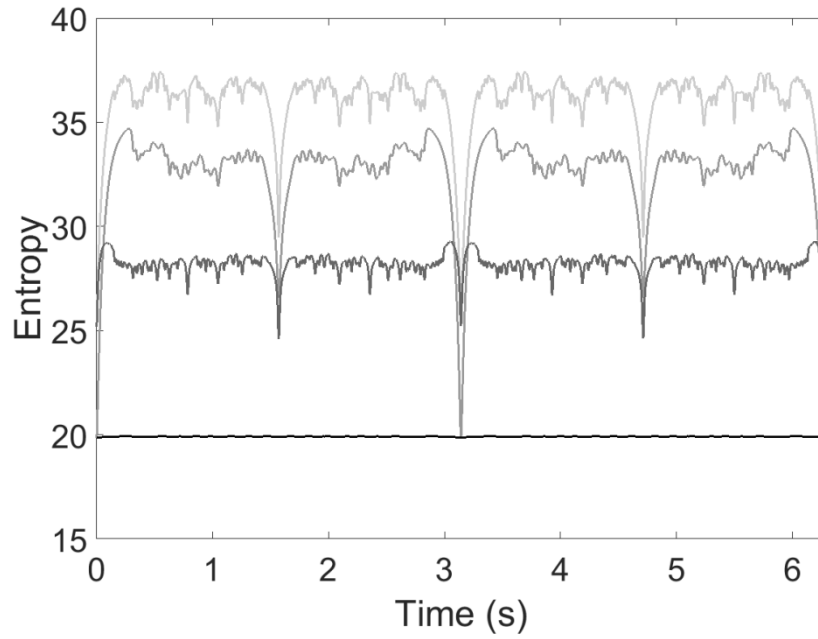


Figure 3 The time evolution of the entropy of a set of 10 uncoupled oscillators. The four curves, from bottom to top (in decreasing darkness) are: (i) the Shannon entropy, (ii) the coarse grained Shannon entropy, (iii) the Boltzmann-type entropy, (iv) the coarse grained Boltzmann-type entropy.

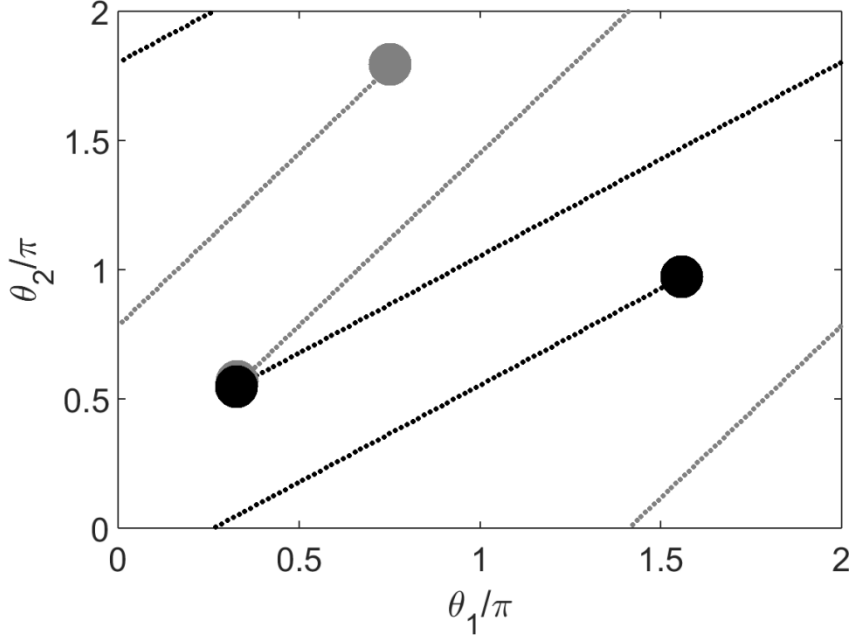


Figure 4 The time evolution of a cloud of angle variables for a two degree of freedom nonlinear integrable system. Two clouds are shown, corresponding to different values of the action variables. The clouds are initially in close proximity, but move apart with increasing time since they have different angular rates. Within each cloud there is no dispersion, since all points have the same angular rates. Note that when a cloud reaches an edge of the domain then the position jumps by 2π . For many values of the action variables the clouds will fill the domain, and coarse grain averaging over the action variables will lead to a uniform distribution in the angle variables.

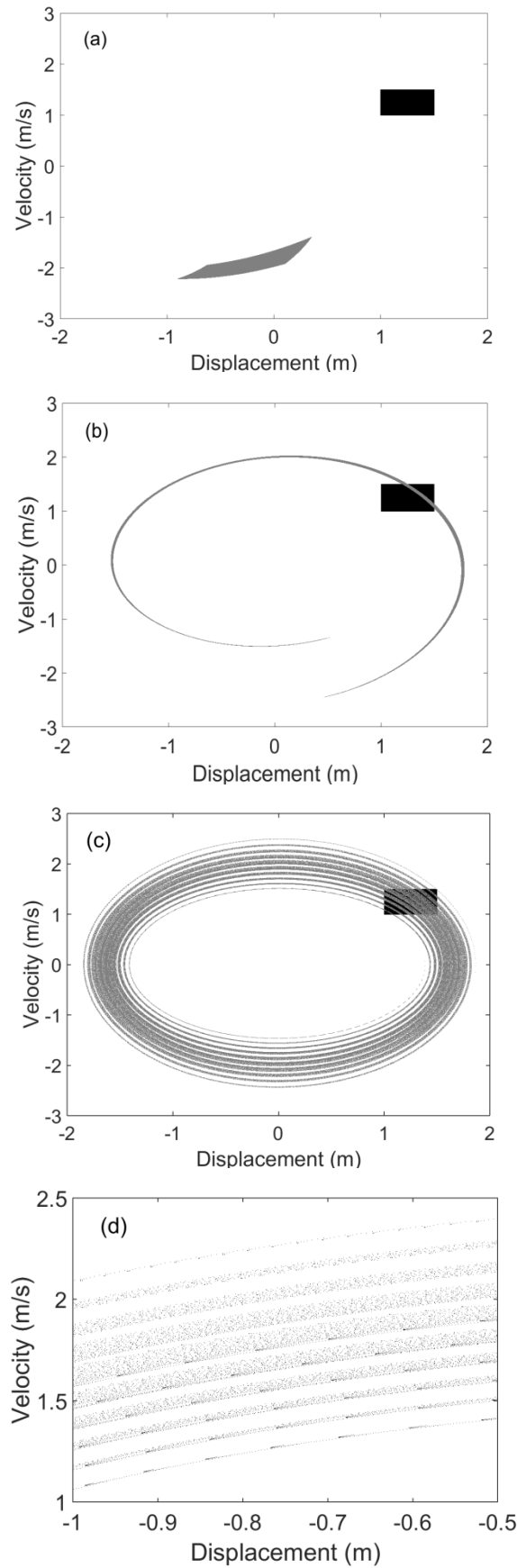


Figure 5. The time evolution of the probability density function of a Duffing oscillator. The dark rectangle in Figures (a) to (c) represents the starting jpdf at $t=0$. The subplots represent: (a) $t=2s$, (b) $t=20s$, (c) $t=200s$, (d) a close up of the response for $t=200s$ showing that the spiral has 10 loops at this time.

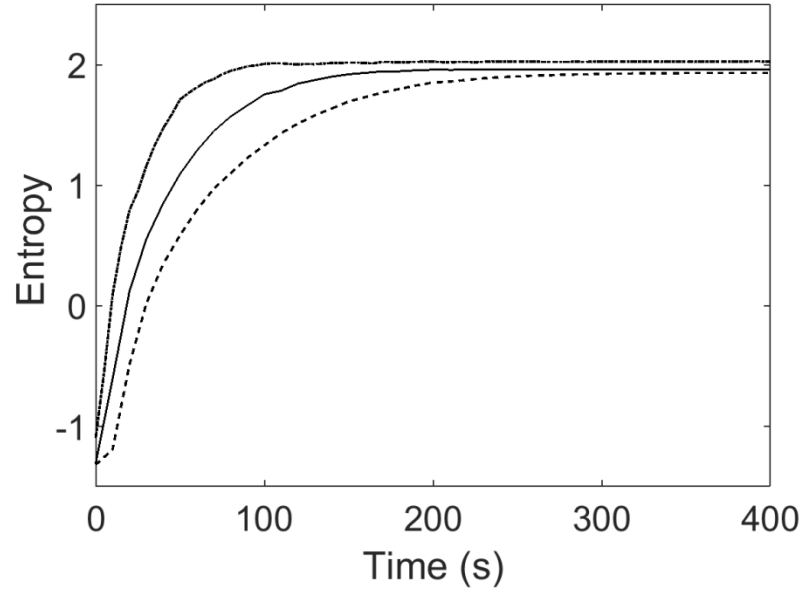


Figure 6 The time evolution of the coarse grained entropy of the Duffing oscillator for various (square) cell sizes. The cell side length associated with the three curves, from the bottom curve to the top curve, are respectively 0.05, 0.1, and 0.2.

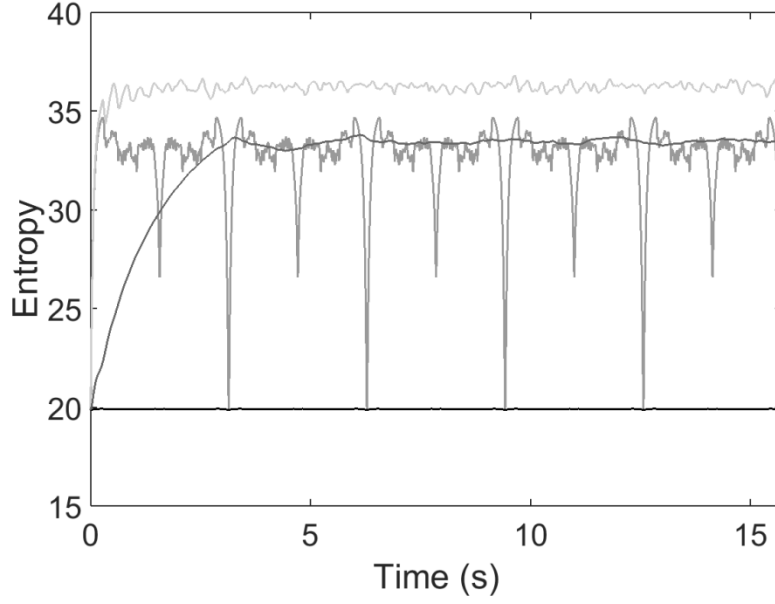


Figure 7 The time evolution of the entropy of a set of 10 uncoupled oscillators. The four curves, from bottom to top (in decreasing darkness) are: (i) the Shannon entropy of a deterministic system, (ii) the Shannon entropy of a random system, (iii) the Boltzmann-type entropy of a deterministic system, (iv) the Boltzmann-type entropy of a random system.